

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Barenkamp, Stephen J
- (ii) TITLE OF INVENTION: High Molecular Weight Surface Proteins of Non-Typeable Haemophilus
- (iii) NUMBER OF SEQUENCES: 11
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Shoemaker and Mattare, Ltd.
  - (B) STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza Bldg. 1
  - (C) CITY: Arlington
  - (D) STATE: Virginia
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 22202-0286
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/617,697
  - (B) FILING DATE: 01-APR-1996
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/302,832
  - (B) FILING DATE: 05-OCT-1994
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US PCT/US93/02166
  - (B) FILING DATE: 16-MAR-1993
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Berkstresser, Jerry W
  - (B) REGISTRATION NUMBER: 22,651
  - (C) REFERENCE/DOCKET NUMBER: 1038-557
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (703) 415-0810
  - (B) TELEFAX: (703) 415-0813

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5116 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTTT GCAGTCTATA TGCAAATATT TTAATAAATA GTATAAATCC GCCATATAAA	120
ATGGTATAAT CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTC ATCTTTTCATC	180

TTTCATCTTT CATCTTTCAT CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTC	240
ACATGCCCTG ATGAACCGAG GGAAGGGAGG GAGGGGCAAG AATGAAGAGG GAGCTGAACG	300
AACGCAAATG ATAAAGTAAT TTAATTGTTC AACTAACCTT AGGAGAAAAT ATGAACAAGC	360
TATATCGTCT CAAATTCAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT GAATTGGCAC	420
GGGGTTGTGA CCATTCCACA GAAAAAGGCA GCGAAAAACC TGCTCGCATG AAAGTGCCTC	480
ACTTAGCGTT AAAGCCACTT TCCGCTATGT TACTATCTTT AGGTGTAACA TCTATTCCAC	540
AATCTGTTTT AGCAAGCGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC	600
AAGTAGATGG TAATAAAACC ATTATCCGCA ACAGTGTGTA CGATATCATT AATTGGAAAC	660
AATTTAACAT CGACCAAAT GAAATGGTGC AGTTTTTACA AGAAAACAAC AACTCCGCCG	720
TATTCAACCG TGTTACATCT AACCAAATCT CCAATTAAA AGGGATTTTA GATTCTAACG	780
GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA ATTATTAACA	840
CTAATGGCTT TACGGCTTCT ACGCTAGACA TTTCTAACGA AACATCAAG GCGCGTAATT	900
TCACCTTCGA GCAAACCAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC GGTTTAATTA	960
CTGTCCGTAA AGACGGCAGT GTAAATCTTA TTGGTGGCAA AGTGAAAAAC GAGGGTGTGA	1020
TTAGCGTAAA TGGTGGCAGC ATTTCTTTAC TCGCAGGGCA AAAAATCACC ATCAGCGATA	1080
TAATAAACC AACCATTACT TACAGCATTG CCGCGCCTGA AAATGAAGCG GTCAATCTGG	1140
GCGATATTTT TGCCAAAGGC GGTAACATTA ATGTCCGTGC TGCCACTATT CGAAACCAAG	1200
GTAACCTTTC TGCTGATTCT GTAAGCAAAG ATAAAAGCGG CAATATTGTT CTTTCCGCCA	1260
AAGAGGGTGA AGCGGAAATT GGCGGTGTAA TTTCCGCTCA AAATCAGCAA GCTAAAGGCG	1320
GCAAGCTGAT GATTACAGGC GATAAAGTCA CATTAAAAAC AGGTGCAGTT ATCGACCTTT	1380
CAGGTAAAGA AGGGGGAGAA ACTTACCTTG GCGGTGACGA GCGCGGCGAA GGTA AAAAAGG	1440
GCATTCAATT AGCAAAGAAA ACCTCTTTAG AAAAAGGCTC AACCATCAAT GTATCAGGCA	1500
AAGAAAAAGG CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC GGCAATATTA	1560
ACGCTCAAGG TAGTGGTGAT ATCGCTAAAA CCGGTGGTTT TGTGGAGACG TCGGGGCATG	1620
ATTTATTCAT CAAAGACAAT GCAATTGTTG ACGCCAAAGA GTGGTTGTTA GACCCGGATA	1680
ATGTATCTAT TAATGCAGAA ACAGCAGGAC GCAGCAATAC TTCAGAAGAC GATGAATACA	1740
CGGGATCCGG GAATAGTGCC AGCACCCCAA AACGAAACAA AGAAAAGACA ACATTAACAA	1800
ACACAACCTCT TGAGAGTATA CTAAAAAAG GTACCTTTGT TAACATCACT GCTAATCAAC	1860
GCATCTATGT CAATAGCTCC ATTAATTTAT CCAATGGCAG CTTAACTCTT TGGAGTGAGG	1920
GTCGGAGCGG TGGCGGCGTT GAGATTAACA ACGATATTAC CACCGGTGAT GATACCAGAG	1980
GTGCAAACTT AACAAATTAC TCAGGCGGCT GGGTTGATGT TCATAAAAAT ATCTCACTCG	2040
GGGCGCAAGG TAACATAAAC ATTACAGCTA AACAAGATAT CGCCTTTGAG AAAGGAAGCA	2100
ACCAAGTCAT TACAGGTCAA GGGACTATTA CCTCAGGCAA TCAAAAAGGT TTTAGATTTA	2160
ATAATGTCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT CACCACTAAA AGAACCAATA	2220

AATACGCTAT CACAAATAAA TTTGAAGGGA CTTTAAATAT TTCAGGGAAA GTGAACATCT	2280
CAATGGTTTT ACCTAAAAAT GAAAGTGGAT ATGATAAATT CAAAGGACGC ACTTACTGGA	2340
ATTTAACCTC CTTAAATGTT TCCGAGAGTG GCGAGTTTAA CCTCACTATT GACTCCAGAG	2400
GAAGCGATAG TGCAGGCACA CTTACCCAGC CTTATAATTT AAACGGTATA TCATTCAACA	2460
AAGACACTAC CTTTAATGTT GAACGAAATG CAAGAGTCAA CTTTGACATC AAGGCACCAA	2520
TAGGGATAAA TAAGTATTCT AGTTTGAATT ACGCATCATT TAATGGAAAC ATTTCACTTT	2580
CGGGAGGGGG GAGTGTTGAT TTCACACTTC TCGCCTCATC CTCTAACGTC CAAACCCCCG	2640
GTGTAGTTAT AAATTCTAAA TACTTTAATG TTTCAACAGG GTCAAGTTTA AGATTTAAAA	2700
CTTCAGGCTC AACAAAACT GGCTTCTCAA TAGAGAAAGA TTTAACTTTA AATGCCACCG	2760
GAGGCAACAT AACACTTTTG CAAGTTGAAG GCACCGATGG AATGATTGGT AAAGGCATTG	2820
TAGCCAAAAA AAACATAACC TTTGAAGGAG GTAACATCAC CTTTGGCTCC AGGAAAGCCG	2880
TAACAGAAAT CGAAGGCAAT GTTACTATCA ATAACAACGC TAACGTCAC TTTATCGGTT	2940
CGGATTTTGA CAACCATCAA AAACCTTTAA CTATTAAAAA AGATGTCATC ATTAATAGCG	3000
GCAACCTTAC CGCTGGAGGC AATATTGTCA ATATAGCCGG AAATCTTACC GTTGAAAGTA	3060
ACGCTAATTT CAAAGCTATC ACAAATTTCA CTTTTAATGT AGGCGGCTTG TTTGACAACA	3120
AAGGCAATTC AAATATTTC ATTGCCAAAG GAGGGGCTCG CTTTAAAGAC ATTGATAATT	3180
CCAAGAATTT AAGCATCACC ACCAACTCCA GCTCCACTTA CCGCACTATT ATAAGCGGCA	3240
ATATAACCAA TAAAAACGGT GATTTAAATA TTACGAACGA AGGTAGTGAT ACTGAAATGC	3300
AAATTGGCGG CGATGTCTCG CAAAAAGAAG GTAATCTCAC GATTCTTCT GACAAAAATCA	3360
ATATTACCAA ACAGATAACA ATCAAGGCAG GTGTTGATGG GGAGAATTCC GATTCAGACG	3420
CGACAAACAA TGCCAATCTA ACCATTAAAA CCAAAGAATT GAAATTAACG CAAGACCTAA	3480
ATATTTCAAGG TTTCAATAAA GCAGAGATTA CAGCTAAAGA TGGTAGTGAT TTAACATTG	3540
GTAACACCAA TAGTGCTGAT GGTACTAATG CAAAAAAGT AACCTTTAAC CAGGTAAAG	3600
ATTCAAAAAT CTCTGCTGAC GGTCACAAGG TGACACTACA CAGCAAAGTG GAAACATCCG	3660
GTAGTAATAA CAACACTGAA GATAGCAGTG ACAATAATGC CGGCTTAACT ATCGATGCAA	3720
AAAATGTAAC AGTAAACAAC AATATTACTT CTCACAAAGC AGTGAGCATC TCTGCGACAA	3780
GTGGAGAAAT TACCACTAAA ACAGGTACAA CCATTAACGC AACCCTGGT AACGTGGAGA	3840
TAACCGCTCA AACAGGTAGT ATCCTAGGTG GAATTGAGTC CAGCTCTGGC TCTGTAACAC	3900
TTACTGCAAC CGAGGGCGCT CTTGCTGTAA GCAATATTTT GGGCAACACC GTTACTGTTA	3960
CTGCAATAG CGGTGCATTA ACCACTTTGG CAGGCTCTAC AATTAAAGGA ACCGAGAGTG	4020
TAACCACTTC AAGTCAATCA GGCGATATCG GCGGTACGAT TTCTGGTGGC ACAGTAGAGG	4080
TTAAAGCAAC CGAAAGTTTA ACCACTCAAT CCAATTCAAA AATTAAAGCA ACAACAGGCG	4140
AGGCTAACGT AACAAGTGCA ACAGGTACAA TTGGTGGTAC GATTTCCGGT AATACGGTAA	4200
ATGTTACGGC AAACGCTGGC GATTTAACAG TTGGGAATGG CGCAGAAATT AATGCGACAG	4260

AAGGAGCTGC AACCTTAACT ACATCATCGG GCAAATTAAC TACC3AAGCT AGTTCACACA 4320  
 TTACTTCAGC CAAGGGTCAG GTAAATCTTT CAGCTCAGGA TGGTAGCGTT GCAGGAAGTA 4380  
 TTAATGCCGC CAATGTGACA CTAAATACTA CAGGCACTTT AACTACCGTG AAGGGTTCAA 4440  
 ACATTAATGC AACCAGCGGT ACCTTGGTTA TTAACGCAAA AGACGCTGAG CTAAATGGCG 4500  
 CAGCATTGGG TAACCACACA GTGGTAAATG CAACCAACGC AAATGGCTCC GGCAGCGTAA 4560  
 TCGCGACAAC CTCAAGCAGA GTGAACATCA CTGGGGATTT AATCACAATA AATGGATTAA 4620  
 ATATCATTTT AAAAAACGGT ATAAACACCG TACTGTTAAA AGGCGTTAAA ATTGATGTGA 4680  
 AATACATTCA ACCGGGTATA GCAAGCGTAG ATGAAGTAAT TGAAGCGAAA CGCATCCTTG 4740  
 AGAAGGTAAA AGATTTATCT GATGAAGAAA GAGAAGCGTT AGCTAAACTT GGAGTAAGTG 4800  
 CTGTACGTTT TATTGAGCCA AATAATACAA TTACAGTCGA TACACAAAAT GAATTTGCAA 4860  
 CCAGACCATT AAGTCGAATA GTGATTTCTG AAGGCAGGGC GTGTTTCTCA AACAGTGATG 4920  
 GCGCGACGGT GTGCGTTAAT ATCGCTGATA ACGGGCGGTA GCGGTCAGTA ATTGACAAGG 4980  
 TAGATTTTCAT CCTGCAATGA AGTCATTTTA TTTTCGTATT ATTTACTGTG TGGGTAAAG 5040  
 TTCAGTACGG GCTTTACCCA TCTTGTAATA AATTACGGAG AATACAATAA AGTATTTTTA 5100  
 ACAGGTTATT ATTATG 5116

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1536 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu  
 1 5 10 15  
 Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys  
 20 25 30  
 Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys  
 35 40 45  
 Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln  
 50 55 60  
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr  
 65 70 75 80  
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val  
 85 90 95  
 Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
 100 105 110  
 Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val  
 115 120 125

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
 130 135 140  
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
 145 150 155 160  
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
 165 170 175  
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys  
 180 185 190  
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
 195 200 205  
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
 210 215 220  
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
 225 230 235 240  
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
 245 250 255  
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
 260 265 270  
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala  
 275 280 285  
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
 290 295 300  
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
 305 310 315 320  
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
 325 330 335  
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
 340 345 350  
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
 355 360 365  
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
 370 375 380  
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
 385 390 395 400  
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly  
 405 410 415  
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile  
 420 425 430  
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn  
 435 440 445  
 Ala Glu Thr Ala Gly Arg Ser Asn Thr Ser Glu Asp Asp Glu Tyr Thr  
 450 455 460  
 Gly Ser Gly Asn Ser Ala Ser Thr Pro Lys Arg Asn Lys Glu Lys Thr  
 465 470 475 480

Thr Leu Thr Asn Thr Thr Leu Glu Ser Ile Leu Lys Lys Gly Thr Phe  
 485 490 495  
 Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn  
 500 505 510  
 Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly  
 515 520 525  
 Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly  
 530 535 540  
 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn  
 545 550 555 560  
 Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp  
 565 570 575  
 Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr  
 580 585 590  
 Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu  
 595 600 605  
 Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys  
 610 615 620  
 Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys  
 625 630 635 640  
 Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys  
 645 650 655  
 Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu  
 660 665 670  
 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala  
 675 680 685  
 Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys  
 690 695 700  
 Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile  
 705 710 715 720  
 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser  
 725 730 735  
 Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Gly Ser Val Asp Phe Thr  
 740 745 750  
 Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn  
 755 760 765  
 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr  
 770 775 780  
 Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu  
 785 790 795 800  
 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp  
 805 810 815  
 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu  
 820 825 830

Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu  
 835 840 845  
 Gly Asn Val Thr Ile Asn Asn Asn Ala Asn Val Thr Leu Ile Gly Ser  
 850 855 860  
 Asp Phe Asp Asn His Gln Lys Pro Leu Thr Ile Lys Lys Asp Val Ile  
 865 870 875 880  
 Ile Asn Ser Gly Asn Leu Thr Ala Gly Gly Asn Ile Val Asn Ile Ala  
 885 890 895  
 Gly Asn Leu Thr Val Glu Ser Asn Ala Asn Phe Lys Ala Ile Thr Asn  
 900 905 910  
 Phe Thr Phe Asn Val Gly Gly Leu Phe Asp Asn Lys Gly Asn Ser Asn  
 915 920 925  
 Ile Ser Ile Ala Lys Gly Gly Ala Arg Phe Lys Asp Ile Asp Asn Ser  
 930 935 940  
 Lys Asn Leu Ser Ile Thr Thr Asn Ser Ser Ser Thr Tyr Arg Thr Ile  
 945 950 955 960  
 Ile Ser Gly Asn Ile Thr Asn Lys Asn Gly Asp Leu Asn Ile Thr Asn  
 965 970 975  
 Glu Gly Ser Asp Thr Glu Met Gln Ile Gly Gly Asp Val Ser Gln Lys  
 980 985 990  
 Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr Lys Gln  
 995 1000 1005  
 Ile Thr Ile Lys Ala Gly Val Asp Gly Glu Asn Ser Asp Ser Asp Ala  
 1010 1015 1020  
 Thr Asn Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys Leu Thr  
 1025 1030 1035 1040  
 Gln Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala Lys  
 1045 1050 1055  
 Asp Gly Ser Asp Leu Thr Ile Gly Asn Thr Asn Ser Ala Asp Gly Thr  
 1060 1065 1070  
 Asn Ala Lys Lys Val Thr Phe Asn Gln Val Lys Asp Ser Lys Ile Ser  
 1075 1080 1085  
 Ala Asp Gly His Lys Val Thr Leu His Ser Lys Val Glu Thr Ser Gly  
 1090 1095 1100  
 Ser Asn Asn Asn Thr Glu Asp Ser Ser Asp Asn Asn Ala Gly Leu Thr  
 1105 1110 1115 1120  
 Ile Asp Ala Lys Asn Val Thr Val Asn Asn Asn Ile Thr Ser His Lys  
 1125 1130 1135  
 Ala Val Ser Ile Ser Ala Thr Ser Gly Glu Ile Thr Thr Lys Thr Gly  
 1140 1145 1150  
 Thr Thr Ile Asn Ala Thr Thr Gly Asn Val Glu Ile Thr Ala Gln Thr  
 1155 1160 1165  
 Gly Ser Ile Leu Gly Gly Ile Glu Ser Ser Ser Gly Ser Val Thr Leu  
 1170 1175 1180

Thr Ala Thr Glu Gly Ala Leu Ala Val Ser Asn Ile Ser Gly Asn Thr  
 1185 1190 1195 1200  
 Val Thr Val Thr Ala Asn Ser Gly Ala Leu Thr Thr Leu Ala Gly Ser  
 1205 1210 1215  
 Thr Ile Lys Gly Thr Glu Ser Val Thr Thr Ser Ser Gln Ser Gly Asp  
 1220 1225 1230  
 Ile Gly Gly Thr Ile Ser Gly Gly Thr Val Glu Val Lys Ala Thr Glu  
 1235 1240 1245  
 Ser Leu Thr Thr Gln Ser Asn Ser Lys Ile Lys Ala Thr Thr Gly Glu  
 1250 1255 1260  
 Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly Thr Ile Ser Gly  
 1265 1270 1275 1280  
 Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu Thr Val Gly Asn  
 1285 1290 1295  
 Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr Leu Thr Thr Ser  
 1300 1305 1310  
 Ser Gly Lys Leu Thr Thr Glu Ala Ser Ser His Ile Thr Ser Ala Lys  
 1315 1320 1325  
 Gly Gln Val Asn Leu Ser Ala Gln Asp Gly Ser Val Ala Gly Ser Ile  
 1330 1335 1340  
 Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Val  
 1345 1350 1355 1360  
 Lys Gly Ser Asn Ile Asn Ala Thr Ser Gly Thr Leu Val Ile Asn Ala  
 1365 1370 1375  
 Lys Asp Ala Glu Leu Asn Gly Ala Ala Leu Gly Asn His Thr Val Val  
 1380 1385 1390  
 Asn Ala Thr Asn Ala Asn Gly Ser Gly Ser Val Ile Ala Thr Thr Ser  
 1395 1400 1405  
 Ser Arg Val Asn Ile Thr Gly Asp Leu Ile Thr Ile Asn Gly Leu Asn  
 1410 1415 1420  
 Ile Ile Ser Lys Asn Gly Ile Asn Thr Val Leu Leu Lys Gly Val Lys  
 1425 1430 1435 1440  
 Ile Asp Val Lys Tyr Ile Gln Pro Gly Ile Ala Ser Val Asp Glu Val  
 1445 1450 1455  
 Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp Leu Ser Asp Glu  
 1460 1465 1470  
 Glu Arg Glu Ala Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Ile  
 1475 1480 1485  
 Glu Pro Asn Asn Thr Ile Thr Val Asp Thr Gln Asn Glu Phe Ala Thr  
 1490 1495 1500  
 Arg Pro Leu Ser Arg Ile Val Ile Ser Glu Gly Arg Ala Cys Phe Ser  
 1505 1510 1515 1520  
 Asn Ser Asp Gly Ala Thr Val Cys Val Asn Ile Ala Asp Asn Gly Arg  
 1525 1530 1535



## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 4937 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TAAATATACA AGATAATAAA AATAAATCAA GATTTTGTG ATGACAAACA ACAATTACAA	60
CACCTTTTTT GCAGTCTATA TGCAAATATT TTAAAAAAT AGTATAAATC CGCCATATAA	120
AATGGTATAA TCTTTCATCT TTCATCTTTA ATCTTTCATC TTTCATCTTT CATCTTTCAT	180
CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTC ATCTTTCATC TTTCATCTTT	240
CACATGAAAT GATGAACCGA GGAAGGGAG GGAGGGGCAA GAATGAAGAG GGAGCTGAAC	300
GAACGCAAAT GATAAAGTAA TTAATTGTT CAACTAACCT TAGGAGAAAA TATGAACAAG	360
ATATATCGTC TCAAATTCAG CAAACGCCTG AATGCTTTGG TTGCTGTGTC TGAATTGGCA	420
CGGGGTTGTG ACCATTCCAC AGAAAAAGGC TTCCGCTATG TTACTATCTT TAGGTGTAAC	480
CACTTAGCGT TAAAGCCACT TTCCGCTATG TTACTATCTT TAGGTGTAAC ATCTATTCCA	540
CAATCTGTTT TAGCAAGCGG CTTACAAGGA ATGGATGTAG TACACGGCAC AGCCACTATG	600
CAAGTAGATG GTAATAAAAC CATTATCCGC AACAGTGTG ACGCTATCAT TAATTGAAA	660
CAATTTAACA TCGACCAAAA TGAAATGGTG CAGTTTTTAC AAGAAAACAA CAACTCCGCC	720
GTATTCAACC GTGTTACATC TAACCAAATC TCCCAATTAA AAGGGATTTT AGATTCTAAC	780
GGACAAGTCT TTTTAATCAA CCCAAATGGT ATCACAATAG GTAAAGACGC AATTATTAAC	840
ACTAATGGCT TTACGGCTTC TACGCTAGAC ATTTCTAACG AAAACATCAA GGCGCGTAAT	900
TTCACCTTCG AGCAAACCAA AGATAAAGCG CTCGCTGAAA TTGTGAATCA CGGTTTAATT	960
ACTGTGGTA AAGACGGCAG TGTAATCTT ATTGGTGGCA AAGTGAAAAA CGAGGGTGTG	1020
ATTAGCGTAA ATGGTGGCAG CATTTCTTTA CTCGCAGGGC AAAAATCAC CATCAGCGAT	1080
ATAATAAACC CAACCATTAC TTACAGCATT GCCGCGCCTG AAAATGAAGC GGTCAATCTG	1140
GGCGATATTT TTGCCAAAGG CGGTAACATT AATGTCCGTG CTGCCACTAT TCGAAACCAA	1200
GGTAACTTT CTGCTGATTC TGTAAGCAAA GATAAAGCG GCAATATTGT TCTTCCGCC	1260
AAAGAGGGTG AAGCGGAAAT TGGCGGTGTA ATTTCCGCTC AAAATCAGCA AGCTAAAGGC	1320
GGCAAGCTGA TGATTACAGG CGATAAAGTC ACATTAAAAA CAGGTGCAGT TATCGACCTT	1380
TCAGGTAAAG AAGGGGGAGA AACTTACCTT GGCGGTGACG AGCGCGGCGA AGGTAAAAAC	1440
GGCATTCAAT TAGCAAAGAA AACCTCTTTA GAAAAAGGCT CAACCATCAA TGTATCAGGC	1500
AAAGAAAAAG GCGGACGCGC TATTGTGTGG GGCGATATTG CGTTAATTGA CGGCAATATT	1560
AACGCTCAAG GTAGTGGTGA TATCGCTAAA ACCGGTGGTT TTGTGGAGAC ATCGGGGCAT	1620
TATTTATCCA TTGACAGCAA TGCAATTGTT AAAACAAAAG AGTGTTGCT AGACCCTGAT	1680

GATGTAACAA TTGAAGCCGA AGACCCCCTT CGCAATAATA CCGGTATAAA TGATGAATTC	1740
CCAACAGGCA CCGGTGAAGC AAGCGACCCT AAAAAAATA GCGAACTCAA AACAACGCTA	1800
ACCAATACAA CTATTTCAA TTATCTGAAA AACGCCTGGA CAATGAATAT AACGGCATCA	1860
AGAAAACTTA CCGTTAATAG CTCAATCAAC ATCGGAAGCA ACTCCCCTT AATTCTCCAT	1920
AGTAAAGGTC AGCGTGGCGG AGGCGTTCAG ATTGATGGAG ATATTACTTC TAAAGGCGGA	1980
AATTTAACCA TTTATTCTGG CGGATGGGTT GATGTTTATA AAAATATTAC GCTTGATCAG	2040
GGTTTTTTAA ATATTACCGC CGCTTCCGTA GCTTTTGAAG GTGGAAATAA CAAAGCACGC	2100
GACGCGGCAA ATGCTAAAAT TGTCGCCCAG GGCCTGTAA CCATTACAGG AGAGGGAAAA	2160
GATTTACAGG CTAACAACGT ATCTTTAAAC GGAACGGGTA AAGGTCTGAA TATCATTTCA	2220
TCAGTGAATA ATTTAACCCA CAATCTTAGT GGCACAATTA ACATATCTGG GAATATAACA	2280
ATTAACCAAA CTACGAGAAA GAACACCTCG TATTGGCAA CCAGCCATGA TTCGCACTGG	2340
AACGTCAGTG CTCTTAATCT AGAGACAGGC GCAAAATTTA CCTTTATTAA ATACATTTCA	2400
AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG CAGGGGTGAA TTTTAACGGC	2460
GTAAATGGCA ACATGTCATT CAATCTCAA GAAGGAGCGA AAGTTAATTT CAAATTAATA	2520
CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTC GGTTTTTAGC CAATATCACA	2580
GCCACTGGTG GGGGCTCTGT TTTTTTTGAT ATATATGCCA ACCATTCTGG CAGAGGGGCT	2640
GAGTTAAAAA TGAGTGAAAT TAATATCTCT AACGGCGCTA ATTTTACCTT AAATTCCTAT	2700
GTTCGCGGCG ATGACGCTTT TAAATCAAC AAAGACTTAA CCATAAATGC AACCAATTCA	2760
AATTTACAGC TCAGACAGAC GAAAGATGAT TTTTATGACG GGTACGCACG CAATGCCATC	2820
AATTCAACCT ACAACATATC CATTCTGGGC GGTAATGTCA CCCTTGGTGG ACAAACCTCA	2880
AGCAGCAGCA TTACGGGGAA TATTACTATC GAGAAAGCAG CAAATGTTAC GCTAGAAGCC	2940
AATAACGCCC CTAATCAGCA AAACATAAGG GATAGAGTTA TAAACCTGG CAGCTTGCTC	3000
GTTAATGGGA GTTTAAGTTT AACTGGCGAA AATGCAGATA TTAAAGGCAA TCTCACTATT	3060
TCAGAAAGCG CCACTTTTAA AGGAAAGACT AGAGATACCC TAAATATCAC CGGCAATTTT	3120
ACCAATAATG GCACTGCCGA AATTAATATA ACACAAGGAG TGGTAAACT TGGCAATGTT	3180
ACCAATGATG GTGATTTAAA CATTACCACT CACGCTAAAC GCAACCAAAG AAGCATCATC	3240
GGCGGAGATA TAATCAACAA AAAAGGAAGC TTAAATATTA CAGACAGTAA TAATGATGCT	3300
GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA ACCTCACGAT TTCTTCCGAT	3360
AAAATTAATA TCACCAAACA GATAACAATC AAAAAGGGTA TTGATGGAGA GGAATCTAGT	3420
TCAGATGCGA CAAGTAATGC CAACCTAACT ATTTAAACCA AAGAATTGAA ATTGACAGAA	3480
GACCTAAGTA TTTACAGTTT CAATAAAGCA GAGATTACAG CCAAAGATGG TAGAGATTTA	3540
ACTATTGGCA ACAGTAATGA CGGTAACAGC GGTGCCGAAG CCAAACAGT AACTTTTAAC	3600
AATGTTAAAG ATTCAAAAAT CTCTGCTGAC GGTCACAATG TGACACTAAA TAGCAAAGTG	3660
AAAACATCTA GCAGCAATGG CGGACGTGAA AGCAATAGCG ACAACGATAC CGGCTTAACT	3720

ATTACTGCAA AAAATGTAGA AGTAAACAAA GATATTACTT CTCTCAAAAC AGTAAATATC 3780  
 ACCGCGTCGG AAAAGGTTAC CACCACAGCA GGCTCGACCA TTAACGCAAC AAATGGCAAA 3840  
 GCAAGTATTA CAACCAAAAC AGGTGATATC AGCGGTACGA TTTCCGGTAA CACGGTAAGT 3900  
 GTTAGCGCGA CTGGTGATTT AACCACTAAA TCCGGCTCAA AAATTGAAGC GAAATCGGGT 3960  
 GAGGCTAATG TAACAAGTGC AACAGGTACA ATTGGCGGTA CAATTTCCGG TAATACGGTA 4020  
 AATGTTACGG CAAACGCTGG CGATTTAACA GTTGGGAATG GCGCAGAAAT TAATGCGACA 4080  
 GAAGGAGCTG CAACCTTAAC CGCAACAGGG AATACCTTGA CTACTGAAGC CGGTTCTAGC 4140  
 ATCACTTCAA CTAAGGGTCA GGTAGACCTC TTGGCTCAGA ATGGTAGCAT CGCAGGAAGC 4200  
 ATTAATGCTG CTAATGTGAC ATTAAATACT ACAGGCACCT TAACCACCGT GGCAGGCTCG 4260  
 GATATTAAAG CAACCAGCGG CACCTTGGTT ATTAACGCAA AAGATGCTAA GCTAAATGGT 4320  
 GATGCATCAG GTGATAGTAC AGAAGTGAAT GCAGTCAACG CAAGCGGCTC TGGTAGTGTG 4380  
 ACTGCGGCAA CCTCAAGCAG TGTGAATATC ACTGGGGATT TAAACACAGT AAATGGGTTA 4440  
 AATATCATTT CGAAAGATGG TAGAAACACT GTGCGCTTAA GAGGCAAGGA AATTGAGGTG 4500  
 AAATATATCC AGCCAGGTGT AGCAAGTGTA GAAGAAGTAA TTGAAGCGAA ACGCGTCCTT 4560  
 GAAAAAGTAA AAGATTTATC TGATGAAGAA AGAGAAACAT TAGCTAAACT TGGTGTAAGT 4620  
 GCTGTACGTT TTGTTGAGCC AAATAATACA ATTACAGTCA ATACACAAA TGAATTTACA 4680  
 ACCAGACCGT CAAGTCAAGT GATAATTTCT GAAGGTAAGG CGTGTCTCTC AAGTGGTAAT 4740  
 GGCGCACGAG TATGTACCAA TGTGCTGAC GATGGACAGC CGTAGTCAGT AATTGACAAG 4800  
 GTAGATTTCA TCCTGCAATG AAGTCATTTT ATTTTCGTAT TATTTACTGT GTGGGTAAAA 4860  
 GTTCAGTACG GGCTTTACCC ATCTTGTAAG AAATTACGGA GAATACAATA AAGTATTTTT 4920  
 AACAGGTTAT TATTATG 4937

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1477 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu  
 1 5 10 15  
 Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys  
 20 25 30  
 Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys  
 35 40 45  
 Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln  
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr  
 65 70 75 80  
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val  
 85 90 95  
 Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
 100 105 110  
 Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val  
 115 120 125  
 Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
 130 135 140  
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
 145 150 155 160  
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
 165 170 175  
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys  
 180 185 190  
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
 195 200 205  
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
 210 215 220  
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
 225 230 235 240  
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
 245 250 255  
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
 260 265 270  
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala  
 275 280 285  
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
 290 295 300  
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
 305 310 315 320  
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
 325 330 335  
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
 340 345 350  
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
 355 360 365  
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
 370 375 380  
 Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
 385 390 395 400  
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly  
 405 410 415

Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile  
 420 425 430  
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn  
 435 440 445  
 Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro  
 450 455 460  
 Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys  
 465 470 475 480  
 Thr Thr Leu Thr Asn Thr Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp  
 485 490 495  
 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile  
 500 505 510  
 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg  
 515 520 525  
 Gly Gly Gly Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Gly Asn  
 530 535 540  
 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr  
 545 550 555 560  
 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu  
 565 570 575  
 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala  
 580 585 590  
 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn  
 595 600 605  
 Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser  
 610 615 620  
 Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly  
 625 630 635 640  
 Asn Ile Thr Ile Asn Gln Thr Thr Arg Lys Asn Thr Ser Tyr Trp Gln  
 645 650 655  
 Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr  
 660 665 670  
 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly  
 675 680 685  
 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val  
 690 695 700  
 Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Gly Ala Lys Val Asn Phe  
 705 710 715 720  
 Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile  
 725 730 735  
 Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Gly Ser Val Phe Phe  
 740 745 750  
 Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser  
 755 760 765

76.

Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val  
 770 775 780  
 Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala  
 785 790 795 800  
 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp  
 805 810 815  
 Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu  
 820 825 830  
 Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ser Ile Thr  
 835 840 845  
 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn  
 850 855 860  
 Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly  
 865 870 875 880  
 Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp  
 885 890 895  
 Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys  
 900 905 910  
 Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr  
 915 920 925  
 Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr  
 930 935 940  
 Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg  
 945 950 955 960  
 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile  
 965 970 975  
 Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser  
 980 985 990  
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr  
 995 1000 1005  
 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser  
 1010 1015 1020  
 Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys  
 1025 1030 1035 1040  
 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr  
 1045 1050 1055  
 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn  
 1060 1065 1070  
 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser  
 1075 1080 1085  
 Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys  
 1090 1095 1100  
 Thr Ser Ser Ser Asn Gly Gly Arg Glu Ser Asn Ser Asp Asn Asp Thr  
 1105 1110 1115 1120

Gly Leu Thr Ile Thr Ala Lys Asn Val Glu Val Asn Lys Asp Ile Thr  
 1125 1130 1135  
 Ser Leu Lys Thr Val Asn Ile Thr Ala Ser Glu Lys Val Thr Thr Thr  
 1140 1145 1150  
 Ala Gly Ser Thr Ile Asn Ala Thr Asn Gly Lys Ala Ser Ile Thr Thr  
 1155 1160 1165  
 Lys Thr Gly Asp Ile Ser Gly Thr Ile Ser Gly Asn Thr Val Ser Val  
 1170 1175 1180  
 Ser Ala Thr Val Asp Leu Thr Thr Lys Ser Gly Ser Lys Ile Glu Ala  
 1185 1190 1195 1200  
 Lys Ser Gly Glu Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly  
 1205 1210 1215  
 Thr Ile Ser Gly Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu  
 1220 1225 1230  
 Thr Val Gly Asn Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr  
 1235 1240 1245  
 Leu Thr Ala Thr Gly Asn Thr Leu Thr Thr Glu Ala Gly Ser Ser Ile  
 1250 1255 1260  
 Thr Ser Thr Lys Gly Gln Val Asp Leu Leu Ala Gln Asn Gly Ser Ile  
 1265 1270 1275 1280  
 Ala Gly Ser Ile Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr  
 1285 1290 1295  
 Leu Thr Thr Val Ala Gly Ser Asp Ile Lys Ala Thr Ser Gly Thr Leu  
 1300 1305 1310  
 Val Ile Asn Ala Lys Asp Ala Lys Leu Asn Gly Asp Ala Ser Gly Asp  
 1315 1320 1325  
 Ser Thr Glu Val Asn Ala Val Asn Ala Ser Gly Ser Gly Ser Val Thr  
 1330 1335 1340  
 Ala Ala Thr Ser Ser Ser Val Asn Ile Thr Gly Asp Leu Asn Thr Val  
 1345 1350 1355 1360  
 Asn Gly Leu Asn Ile Ile Ser Lys Asp Gly Arg Asn Thr Val Arg Leu  
 1365 1370 1375  
 Arg Gly Lys Glu Ile Glu Val Lys Tyr Ile Gln Pro Gly Val Ala Ser  
 1380 1385 1390  
 Val Glu Glu Val Ile Glu Ala Lys Arg Val Leu Glu Lys Val Lys Asp  
 1395 1400 1405  
 Leu Ser Asp Glu Glu Arg Glu Thr Leu Ala Lys Leu Gly Val Ser Ala  
 1410 1415 1420  
 Val Arg Phe Val Glu Pro Asn Asn Thr Ile Thr Val Asn Thr Gln Asn  
 1425 1430 1435 1440  
 Glu Phe Thr Thr Arg Pro Ser Ser Gln Val Ile Ile Ser Glu Gly Lys  
 1445 1450 1455  
 Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala  
 1460 1465 1470

Asp Asp Gly Gln Pro  
1475

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAGCGTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTTT GCAGTCTATA TGCAAATATT TTA AAAAATA GTATAAATCC GCCATATAAA	120
ATGGTATAAT CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTC ATCTTTCATC	180
TTTCATCTTT CATCTTTCAT CTTTCATCTT TCATCTTTCA TCTTTCATCT TTCATCTTTC	240
ACATGAAATG ATGAACCGAG GGAAGGGAGG GAGGGGCAAG AATGAAGAGG GAGCTGAACG	300
AACGCAAATG ATAAAGTAAT TTAATTGTTT AACTAACCTT AGGAGAAAAT ATGAACAAGA	360
TATATCGTCT CAAATTCAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT GAATTGGCAC	420
GGGGTTGTGA CCATTCCACA GAAAAAGGCA GCGAAAAACC TGCTCGCATG AAAGTGCCTC	480
ACTTAGCGTT AAAGCCACTT TCCGCTATGT TACTATCTTT AGGTGTAACA TCTATTCCAC	540
AATCTGTTTT AGCAAGCGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC	600
AAGTAGATGG TAATAAAACC ATTATCCGCA ACAGTGTGA CGCTATCATT AATTGGAAAC	660
AATTTAATAT CGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAACAAC AACTCCGCCG	720
TATTCAACCG TGTTACATCT AACCAAATCT CCCAATTAAA AGGGATTTTA GATTCTAACG	780
GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA ATTATTAACA	840
CTAATGGCTT TACGGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG GCGCGTAATT	900
TCACCTTCGA GCAAACCAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC GGTTTAATTA	960
CTGTCGGTAA AGACGGCAGT GTAAATCTTA TTGGTGGCAA AGTGAAAAAC GAGGGTGTGA	1020
TTAGCGTAAA TGGTGGCAGC ATTTCTTTAC TCGCAGGGCA AAAAATCACC ATCAGCGATA	1080
TAATAAACCC AACCATTACT TACAGCATTG CCGCGCCTGA AAATGAAGCG GTCAATCTGG	1140
GCGATATTTT TGCCAAAGGC GGTAACATTA ATGTCCGTGC TGCCACTATT CGAAACCAAG	1200
CTTTCCGCCA AAGAGGGTGA AGCGGAAATT GGCGGTGTAA TTTCCGCTCA AAATCAGCAA	1260
GCTAAAGGCG GCAAGCTGAT GATTACAGGC GATAAAGTCA CATTAAAAAC AGGTGCAGTT	1320
ATCGACCTTT CAGGTAAAGA AGGGGGAGAA ACTTACCTTG GCGGTGACGA GCGCGGCGAA	1380
GGTAAAAACG GCATTCAATT AGCAAAGAAA ACCTCTTTAG AAAAAGGCTC AACCATCAAT	1440
GTATCAGGCA AAGAAAAAGG CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC	1500



GGCAATATTA ACGCTCAAGG TAGTGGTGAT ATCGCTAAAA CCGGTGGTTT TGTGGAGACG	1560
TCGGGGCATG ATTTATTCAT CAAAGACAAT GCAATTGTTG ACGCCAAAGA GTGGTTGTTA	1620
GACCCGATA ATGTATCTAT TAATGCAGAA ACAGCAGGAC GCAGCAATAC TTCAGAAGAC	1680
GATGAATACA CGGGATCCGG GAATAGTGCC AGCACCCCAA AACGAAACAA AGAAAAGACA	1740
ACATTAACAA ACACAACCTCT TGAGAGTATA CTAAAAAAG GTACCTTTGT TAACATCACT	1800
GCTAATCAAC GCATCTATGT CAATAGCTCC ATTAATTTAT CCAATGGCAG CTTAACTCTT	1860
TGGAGTGAGG GTCGGAGCGG TGGCGGCGTT GAGATTAACA ACGATATTAC CACCGGTGAT	1920
GATACCAGAG GTGCAAACTT AACAAATTAC TCAGGCGGCT GGGTTGATGT TCATAAAAAT	1980
ATCTCACTCG GGGCGCAAGG TAACATAAAC ATTACAGCTA AACAGATAT CGCCTTTGAG	2040
AAAGGAAGCA ACCAAGTCAT TACAGGTCAA GGGACTATTA CCTCAGGCAA TCAAAAAGGT	2100
TTTAGATTTA ATAATGTCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT CACCACTAAA	2160
AGAACCAATA AATACGCTAT CACAAATAAA TTTGAAGGGA CTTTAAATAT TTCAGGGAAA	2220
GTGAACATCT CAATGGTTTT ACCTAAAAAT GAAAGTGGAT ATGATAAATT CAAAGGACGC	2280
ACTTACTGGA ATTTAACCTC GAAAGTGGAT ATGATAAATT CAAAGGACGC CCTCACTATT	2340
GACTCCAGAG GAAGCGATAG TGCAGGCACA CTTACCCAGC CTTATAATTT AAACGGTATA	2400
TCATTCAACA AAGACACTAC CTTTAATGTT GAACGAAATG CAAGAGTCAA CTTTGACATC	2460
AAGGCACCAA TAGGGATAAA TAAGTATTCT AGTTTGAATT ACGCATCATT TAATGGAAAC	2520
ATTCAGTTT CGGGAGGGGG GAGTGTGAT TTCACACTTC TCGCCTCATC CTCTAACGTC	2580
CAAAACCCCG GTGTAGTTAT AAATTCTAAA TACTTTAATG TTTCAACAGG GTCAAGTTTA	2640
AGATTTAAAA CTTCAGGCTC AACAAAACT GGCTTCTCAA TAGAGAAAGA TTTAACTTTA	2700
AATGCCACCG GAGGCAACAT AACACTTTTG CAAGTTGAAG GCACCGATGG AATGATTGGT	2760
AAAGGCATTG TAGCCAAAAA AAACATAACC TTTGAAGGAG GTAAGATGAG GTTTGGCTCC	2820
AGGAAAGCCG TAACAGAAAT CGAAGGCAAT GTTACTATCA ATAACAACGC TAACGTCACT	2880
CTTATCGGTT CGGATTTTGA CAACCATCAA AAACCTTTAA CTATTAAAAA AGATGTCATC	2940
ATTAATAGCG GCAACCTTAC CGCTGGAGGC AATATTGTCA ATATAGCCGG AAATCTTACC	3000
GTTGAAAGTA ACGCTAATTT CAAAGCTATC ACAAATTTCA CTTTAAATGT AGGCGGCTTG	3060
TTTGACAACA AAGGCAATTC AAATATTTCC ATTGCCAAAG GAGGGGCTCG CTTTAAAGAC	3120
ATTGATAATT CCAAGAATTT AAGCATCACC ACCAACTCCA GCTCCACTTA CCGCACTATT	3180
ATAAGCGGCA ATATAACCAA TAAAAACGGT GATTTAAATA TTACGAACGA AGGTAGTGAT	3240
ACTGAAATGC AAATTGGCGG CGATGTCTCG CAAAAGAAG GTAATCTCAC GATTCTTCT	3300
GACAAAATCA ATATTACCAA ACAGATAACA ATCAAGGCAG GTGTTGATGG GGAGAATTCC	3360
GATTCAAGACG CGACAAACAA TGCCAATCTA ACCATTAAAA CCAAAGAATT GAAATTAACG	3420
CAAGACCTAA ATATTTCAAG TTTCAATAAA GCAGAGATTA CAGCTAAAGA TGGTAGTGAT	3480
TTAACTATTG GTAACACCAA TAGTGCTGAT GGTACTAATG CCAAAAAAGT AACCTTTAAC	3540

CAGGTTAAAG ATTCAAAAAT CTCTGCTGAC GGTCAACAAGG TGACACTACA CAGCAAAGTG	3600
GAAACATCCG GTAGTAATAA CAACACTGAA GATAGCAGTG ACAATAATGC CGGCTTAACT	3660
ATCGATGCAA AAAATGTAAC AGTAAACAAC AATATTACTT CTCACAAAGC AGTGAGCATC	3720
TCTGCGACAA GTGGAGAAAT TACCACTAAA ACAGGTACAA CCATTAACGC AACCCTGGT	3780
AACGTGGAGA TAACCGCTCA AACAGGTAGT ATCCTAGGTG GAATTGAGTC CAGCTCTGGC	3840
TCTGTAAACAC TTACTGCAAC CGAGGGCGCT CTTGCTGTAA GCAATATTTC GGGCAACACC	3900
GTTACTGTTA CTGCAAATAG CGGTGCATTA ACCACTTTGG CAGGCTCTAC AATTAAAGGA	3960
ACCGAGAGTG TAACCACTTC AAGTCAATCA GGCGATATCG GCGGTACGAT TTCTGGTGGC	4020
ACAGTAGAGG TTAAAGCAAC CGAAAGTTTA ACCACTCAAT CCAATTCAAA AATTAAAGCA	4080
ACAACAGGCG AGGCTAACGT AACAAGTGCA ACAGGTACAA TTGGTGGTAC GATTTCCGGT	4140
AATACGGTAA ATGTTACGGC AAACGCTGGC GATTTAACAG TTGGGAATGG CGCAGAAATT	4200
AATGCGACAG AAGGAGCTGC AACCTTAACT ACATCATCGG GCAAATTAAC TACCGAAGCT	4260
AGTTCACACA TTACTTCAGC CAAGGGTCAG GTAAATCTTT CAGCTCAGGA TGGTAGCGTT	4320
GCAGGAAGTA TTAATGCCGC CAATGTGACA CTAAATACTA CAGGCACTTT AACTACCGTG	4380
AAGGGTTCAA ACATTAATGC AACCAGCGGT ACCTTGTTA TTAACGCAA AGACGCTGAG	4440
CTAAATGGCG CAGCATTGGG TAACCACACA GTGGTAAATG CAACCAACGC AAATGGCTCC	4500
GGCAGCGTAA TCGCGACAAC CTCAAGCAGA GTGAACATCA CTGGGGATTT AATCACAATA	4560
AATGGATTAA ATATCATTTT AAAAAACGGT ATAAACACCG TACTGTTAAA AGGCGTTAAA	4620
ATTGATGTGA AATACATTCA ACCGGGTATA GCAAGCGTAG ATGAAGTAAT TGAAGCGAAA	4680
CGCATCCTTG AGAAGGTAAA AGATTTATCT GATGAAGAAA GAGAAGCGTT AGCTAAACTT	4740
GGCGTAAGTG CTGTACGTTT TATTGAGCCA AATAATACAA TTACAGTCGA TACACAAAAT	4800
GAATTTGCAA CCAGACCATT AAGTCGAATA GTGATTTCTG AAGGCAGGGC GTGTTTCTCA	4860
AACAGTGATG GCGCGACGGT GTGCGTTAAT ATCGCTGATA ACGGGCGGTA GCGGTCAGTA	4920
ATTGACAAGG TAGATTTTCT CCTGCAATGA AGTCATTTTA TTTTCGTATT ATTTACTGTG	4980
TGGGTAAAG TTCAGTACGG GCTTTACCCA TCTTGTAATA AATTACGGAG AATACAATAA	5040
AGTATTTTTA ACAGGTATT ATTATGAAAA ATATAAAAAG CAGATTAAAA CTCAGTGCAA	5100
TATCAGTATT GCTTGGCCTG GCTTCTTCAT CATGTATGC AGAAGAAGCG TTTTGTAGTA	5160
AAGGCTTTCA GTTATCTGGT GCACTTGAAA CTTTAAGTGA AGACGCCCAA CTGTCTGTAG	5220
CAAAATCTTT ATCTAAATAC CAAGGCTCGC AAACCTTAAC AACCTAAAA ACAGCACAGC	5280
TTGAATTACA GGCTGTGCTA GATAAGATTG AGCCAAATAA GTTTGATGTG ATATTGCCAC	5340
AACAAACCAT TACGGATGGC AATATTATGT TTGAGCTAGT CTCGAAATCA GCCGCAGAAA	5400
GCCAAGTTTT TTATAAGGCG AGCCAGGGTT ATAGTGAAGA AAATATCGCT CGTAGCCTGC	5460
CATCTTTGAA ACAAGGAAAA GTGTATGAAG ATGGTCGTCA GTGGTTCGAT TTGCGTGAAT	5520
TCAATATGGC AAAAGAAAA CCACTTAAAG TCACTCGCGT GCATTACGAG TTAAACCCTA	5580

AAAACAAAAC CTCTGATTTG GTAGTTGCAG GTTTTTCGCC TTTTGGCAAA ACGCGTAGCT	5640
TTGTTTCCTA TGATAATTTT GCGCAAGGG AGTTTAACTA TCAACGTGTA AGTCTAGGTT	5700
TTGTAAATGC CAATTTGACC GGACATGATG ATGTATTAAA TCTAAACGCA TTGACCAATG	5760
TAAAAGCACC ATCAAAATCT TATGCGGTAG GCATAGGATA TACTTATCCG TTTTATGATA	5820
AACACCAATC CTTAAGTCTT TATACCAGCA TGAGTTATGC TGATTCTAAT GATATCGACG	5880
GCTTACCAAG TGCGATTAAT CGTAAATTAT CAAAAGGTCA ATCTATCTCT GCGAATCTGA	5940
AATGGAGTTA TTATCTCCCG ACATTTAACC TTGGAATGGA AGACCAGTTT AAAATTAATT	6000
TAGGCTACAA CTACCGCCAT ATTAATCAAA CATCCGAGTT AAACACCCTG GGTGCAACGA	6060
AGAAAAAATT TGCAGTATCA GCGTAAGTG CAGGCATTGA TGGACATATC CAATTTACCC	6120
CTAAAACAAT CTTTAATATT GATTTAACTC ATCATTATTA CCGAGTAAA TTACCAGGCT	6180
CTTTTGGAAAT GGAGCGCATT GCGGAAACAT TTAATCGCAG CTATCACATT AGCACAGCCA	6240
GTTTAGGGTT GAGTCAAGAG TTTGCTCAAG GTTGGCATT TTAGCAGTCAA TTATCGGGTC	6300
AGTTTACTCT ACAAGATATA AGTAGCATAG ATTTATTCTC TGTAACAGGT ACTTATGGCG	6360
TCAGAGGCTT TAAATACGGC GGTGCAAGTG GTGAGCGCGG TCTTGATGG CGTAATGAAT	6420
TAAGTATGCC AAAATACACC CGCTTTCAAA TCAGCCCTTA TGCGTTTTAT GATGCAGGTC	6480
AGTTCCGTTA TAATAGCGAA AATGCTAAAA CTTACGGCGA AGATATGCAC ACGGTATCCT	6540
CTGCGGGTTT AGGCATTAAA ACCTCTCCTA CACAAAACCT AAGCTTAGAT GCTTTTGTG	6600
CTCGTCGCTT TGCAAATGCC AATAGTGACA ATTTGAATGG CAACAAAAAA CGCACAAGCT	6660
CACCTACAAC CTTCTGGGGT AGATTAAACAT TCAGTTTCTA ACCCTGAAAT TTAATCAACT	6720
GGTAAGCGTT CCGCCTACCA GTTTATAACT ATATGCTTTA CCCGCCAATT TACAGTCTAT	6780
ACGCAACCCT GTTTTCATCC TTATATATCA AACAACTAA GCAAACCAAG CAAACCAAGC	6840
AAACCAAGCA AACCAAGCAA ACCAAGCAAA CCAAGCAAAC CAAGCAAACC AAGCAAACCA	6900
AGCAAACCAA GCAAACCAAG CAAACCAAGC AAACCAAGCA ATGCTAAAAA ACAATTTATA	6960
TGATAAACTA AAACATACTC CATACCATGG CAATACAAGG GATTTAATAA TATGACAAAA	7020
GAAAATTTAC AAAGTGTTC ACAAATACG ACCGCTTCAC TTGTAGAATC AAACAACGAC	7080
CAAACCTCCC TGCAAATACT TAAACAACCA CCCAAACCCA ACCTATTACG CCTGGAACAA	7140
CATGTCGCCA AAAAAGATTA TGAGCTTGCT TGCCGCGAAT TAATGGCGAT TTTGGAAAAA	7200
ATGGACGCTA ATTTTGGAGG CGTTCACGAT ATTGAATTTG ACGCACCTGC TCAGCTGGCA	7260
TATCTACCCG AAAAATACT AATTCATTTT GCCACTCGTC TCGCTAATGC AATTACAACA	7320
CTCTTTTCCG ACCCCGAATT GGCAATTTCC GAAGAAGGGG CATTAAAGAT GATTAGCCTG	7380
CAACGCTGGT TGACGCTGAT TTTTGCCTCT TCCCCCTACG TTAACGCAGA CCAATTCTC	7440
AATAAATATA ATATCAACCC AGATTCCGAA GGTGGCTTTC ATTTAGCAAC AGACAACTCT	7500
TCTATTGCTA AATTCTGTAT TTTTACTTA CCCGAATCCA ATGTCAATAT GAGTTTAGAT	7560
GCGTTATGGG CAGGGAATCA ACAACTTTGT GCTTCATTGT GTTTTGCATT GCAGTCTTCA	7620

82

CGTTTTATTG GTACTGCATC TCGGTTTCAT AAAAGAGCGG TGGTTTTACA GTGGTTTCCT	7680
AAAAAACTCG CCGAAATTGC TAATTTAGAT GAATTGCCTG CAAATATCCT TCATGATGTA	7740
TATATGCACT GCAGTTATGA TTTAGCAAAA AACAAGCACG ATGTTAAGCG TCCATTAAAC	7800
GAACTTGTCC GCAAGCATAT CCTCACGCAA GGATGGCAAG ACCGCTACCT TTACACCTTA	7860
GGTAAAAAGG ACGGCAAACC TGTGATGATG GTACTGCTTG AACATTTTAA TTCGGGACAT	7920
TCGATTTATC GCACGCATTC AACTTCAATG ATTGCTGCTC GAGAAAAATT CTATTTAGTC	7980
GGCTTAGGCC ATGAGGGCGT TGATAACATA GGTCGAGAAG TGTTTGACGA GTTCTTTGAA	8040
ATCAGTAGCA ATAATATAAT GGAGAGACTG TTTTTTATCC GTAAACAGTG CGAAACTTTC	8100
CAACCCGCAG TGTTCTATAT GCCAAGCATT GGCATGGATA TTACCACGAT TTTGTGAGC	8160
AACACTCGGC TTGCCCTAT TCAAGCTGTA GCCTTGGGTC ATCCIGCCAC TACGCATTCT	8220
GAATTTATTG ATTATGTCAT CGTAGAAGAT GATTATGTGG GCAGTGAAGA TTGTTTtagc	8280
GAAACCTTT TAGGCTTACC CAAAGATGCC CTACCTTATG TACCATCTGC ACTCGCCCCA	8340
CAAAAAGTGG ATTATGTACT CAGGGAAAAC CCTGAAGTAG TCAATATCGG TATTGCCGCT	8400
ACCACAATGA AATTAAACCC TGAATTTTTG CTAACATTGC AAGAAATCAG AGATAAAGCT	8460
AAAGTCAAAA TACATTTTCA TTTCGCACTT GGACAATCAA CAGGCTTGAC ACACCTTAT	8520
GTCAAATGGT TTATCGAAAG CTATTTAGGT GACGATGCCA CTGCACATCC CCACGCACCT	8580
TATCACGATT ATCTGGCAAT ATTGCGTGAT TGCGATATGC TACTAAATCC GTTTCCTTTC	8640
GGTAATACTA ACGGCATAAT TGATATGGTT ACATTAGGTT TAGTTGGTGT ATGCAAAACG	8700
GGGGATGAAG TACATGAACA TATTGATGAA GGTCTGTTTA AACGCTTAGG ACTACCAGAA	8760
TGGCTGATAG CCGACACACG AGAAACATAT ATTGAATGTG CTTTGCCTCT AGCAGAAAAC	8820
CATCAAGAAC GCCTTGAAC TCGTCGTTAC ATCATAGAAA ACAACGGCTT ACAAAGCTT	8880
TTTACAGGCG ACCCTCGTCC ATTGGGCAAA ATACTGCTTA AGAAAACAAA TGAATGGAAG	8940
CGGAAGCACT TGAGTAAAAA ATAACGGTTT TTAAAGTAA AAGTGCGGTT AATTTTCAAA	9000
GCGTTTTAAA AACCTCTCAA AAATCAACCG CACTTTTATC TTTATAACGC TCCCGCGCGC	9060
TGACAGTTTA TCTCTTCTT AAAATACCCA TAAATTTGTG GCAATAGTTG GGTAATCAAA	9120
TTCAATTGTT GATACGGCAA ACTAAAGACG GCGCGTTCTT CGGCAGTCAT C	9171

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCCACTTCA ATTTTGGATT GTTGAAATTC AACTAACCAA AAAGTGCGGT TAAATCTGT	60
--	----

GGAGAAAATA GGTGTAGTG AAGAACGAGG TAATTGTTCA AAAGGATAAA GCTCTCTTAA	120
TTGGGCATTG GTTGGCGTTT CTTTTTCGGT TAATAGTAAA TTATATTCTG GACGACTATG	180
CAATCCACCA ACAACTTTAC CGTTGGTTTT AAGCGTTAAT GTAAGTTCTT GCTCTTCTTG	240
GCGAATACGT AATCCCATTT TTTGTTTAGC AAGAAAATGA TCGGGATAAT CATAATAGGT	300
GTTGCCCAAA AATAAATTTT GATGTTCTAA AATCATAAAT TTTGCAAGAT ATTGTGGCAA	360
TTCAATACCT ATTTGTGGCG AAATCGCCAA TTTAATTCA ATTTCTTGTA GCATAATATT	420
TCCCACTCAA ATCAACTGGT TAAATATACA AGATAATAAA AATAAATCAA GATTTTTGTG	480
ATGACAAACA ACAATTACAA CACCTTTTTT GCAGTCTATA TGCAAATATT TTAAAAAAT	540
AGTATAAATC CGCCATATAA AATGGTATAA TCTTTCATCT TTCATCTTC ATCTTTCATC	600
TTTCATCTTT CATCTTTCAT CTTTCATCTT TCATCTTCA TCTTTCATCT TTCATCTTC	660
ATCTTTCATC TTTCATCTTT CACATGAAAT GATGAACCGA GGGAAGGGAG GGAGGGGCAA	720
GAATGAAGAG GGAGCTGAAC GAACGCAAAT GATAAAGTAA TTTAATTGTT CAACTAACCT	780
TAGGAGAAAA TATGAACAAG ATATATCGTC TCAAATTCAG CAAACGCCTG AATGCTTTGG	840
TTGCTGTGTC TGAATTGGCA CGGGGTTGTG ACCATTCCAC AGAAAAAGGC AGCGAAAAAC	900
CTGCTCGCAT GAAAGTGCCT CACTTAGCGT TAAAGCCACT TTCCGCTATG TTACTATCTT	960
TAGGTGTAAC ATCTATTCCA CAATCTGTTT TAGCAAGCGG CAATTTAACA TCGACCAAAA	1020
TGAAATGGTG CAGTTTTTAC AAGAAAACAA GTAATAAAAC CATTATCCGC AACAGTGTG	1080
ACGCTATCAT TAATTGGAAA CAATTTAACA TCGACCAAAA TGAAATGGTG CAGTTTTTAC	1140
AAGAAAACAA CAACTCCGCC GTATTCAACC GTGTTACATC TAACCAAATC TCCCAATTAA	1200
AAGGGATTTT AGATTCTAAC GGACAAGTCT TTTAATCAA CCCAAATGGT ATCACAATAG	1260
GTAAAGACGC AATTATTAAC ACTAATGGCT TTACGGCTTC TACGCTAGAC ATTTCTAACG	1320
AAAACATCAA GCGCGTAAT TTCACCTTCG AGCAAACCAA AGATAAAGCG CTCGCTGAAA	1380
TTGTGAATCA CGGTTTAATT ACTGTCGGTA AAGACGGCAG TGTAATCTT ATTGGTGGCA	1440
AAGTGAAAAA CGAGGGTGTG ATTAGCGTAA ATGGTGGCAG CATTTCTTTA CTCGCAGGGC	1500
AAAAATCAC CATCAGCGAT ATAATAAACC CAACCATTAC TTACAGCATT GCCGCGCCTG	1560
AAAATGAAGC GGTCAATCTG GCGGATATTT TTGCCAAAGG CGGTAACATT AATGTCCGTG	1620
CTGCCACTAT TCGAAACCAA GGTAACTTT CTGCTGATTC TGTAAGCAAA GATAAAGCG	1680
GCAATATTGT TCTTCCGCC AAAGAGGGTG AAGCGGAAAT TGGCGGTGTA ATTTCCGCTC	1740
AAAATCAGCA AGCTAAAGGC GGCAAGCTGA TGATAAAGTC CGATAAAGTC ACATTAAAAA	1800
CAGGTGCAGT TATCGACCTT TCAGGTAAAG AAGGGGGAGA AACTTACCTT GCGGGTGACG	1860
AGCGCGGCGA AGGTAAAAAC GGCATTCAAT TAGCAAGAA AACCTCTTTA GAAAAAGGCT	1920
CAACCATCAA TGTATCAGGC AAAGAAAAAG GCGGACGCGC TATTGTGTGG GGCGATATTG	1980
CGTTAATTGA CGGCAATATT AACGCTCAAG GTAGTGGTGA TATCGCTAAA ACCGGTGGTT	2040
TTGTGGAGAC ATCGGGGCAT TATTTATCCA TTGACAGCAA TGCAATTGTT AAAACAAAAG	2100

AGTGGTTGCT AGACCCTGAT GATGTAACAA TTGAAGCCGA AGACCCCCTT CGCAATAATA	2160
CCGGTATAAA TGATGAATTC CCAACAGGCA CCGGTGAAGC AAGCGACCCT AAAAAAATA	2220
GCGAACTCAA AACAAAGCTA ACCAATACAA CTATTTCAA TTATCTGAAA AACGCCTGGA	2280
CAATGAATAT AACGGCATCA AGAAAACTTA CCGTTAATAG CTCAATCAAC ATCGGAAGCA	2340
ACTCCCCTT AATTCTCCAT AGTAAAGGTC AGCGTGGCGG AGGCGTTCAG ATTGATGGAG	2400
ATATTACTTC TAAAGGCGGA AATTTAACCA TTTATTCTGG CGGATGGGTT GATGTTTATA	2460
AAAATATTAC GCTTGATCAG GGTTTTTTAA ATATTACCGC CGCTCCGTA GCTTTTGAAG	2520
GTGGAAATAA CAAAGCACGC GACGCGGCAA ATGTAAAAT TGTCGCCCAG GGCAGTGTA	2580
CCATTACAGG AGAGGGAAAA GATTTAGGG CTAACAACGT ATCTTTAAAC GGAACGGGTA	2640
AAGGTCTGAA TATCATTTCA TCAGTGAATA ATTTAACCCA CAATCTTAGT GGCACAATTA	2700
ACATATCTGG GAATATAACA ATTAACCAA CTACGAGAAA GAACACCTCG TATTGGCAA	2760
CCAGCCATGA TTCGCACTGG AACGTCAGTG CTCTTAATCT AGAGACAGGC GCAAATTTTA	2820
CCTTTATTAA ATACATTTCA AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG	2880
CAGGGGTGAA TTTTAACGGC GTAAATGGCA ACATGTCATT CAATCTCAA GAAGGAGCGA	2940
AAGTTAATTT CAAATTAAAA CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTC	3000
GGTTTTTAGC CAATATCACA GCCACTGGTG GGGGCTCTGT TTTTTTTGAT ATATATGCCA	3060
ACCATTCTGG CAGAGGGGCT GAGTTAAAA TGAGTGAAAT TAATATCTCT AACGGCGCTA	3120
ATTTTACCTT AAATTCCCAT GTTCGCGCG ATGACGCTTT TAAAATCAAC AAAGACTTAA	3180
CCATAAATGC AACCAATTCA AATTCAGCC TCAGACAGAC GAAAGATGAT TTTTATGACG	3240
GGTACGCACG CAATGCCATC AATTCAACCT ACAACATATC CATTCTGGGC GGTAATGTCA	3300
CCCTTGGTGG ACAAACCTCA AGCAGCAGCA TTACGGGGAA TATTACTATC GAGAAAGCAG	3360
CAAATGTTAC GCTAGAAGCC AATAACGCCC CTAATCAGCA AAACATAAGG GATAGAGTTA	3420
TAAAACTTGG CAGCTTGCTC GTTAATGGGA GTTTAAGTTT AACTGGCGAA AATGCAGATA	3480
TTAAAGGCAA TCTCACTATT TCAGAAAGCG CCACTTTTAA AGGAAAGACT AGAGATACCC	3540
TAAATATCAC CGGCAATTTT ACCAATAATG GCACTGCCGA AATTAATATA ACACAAGGAG	3600
TGGTAAAACT TGGCAATGTT ACCAATGATG GTGATTTAAA CATTACCACT CACGCTAAAC	3660
GCAACCAAAG AAGCATCATC GGCGGAGATA TAATCAACAA AAAAGGAAGC TTAAATATTA	3720
CAGACAGTAA TAATGATGCT GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA	3780
ACCTCACGAT TTCTCCGAT AAAATTAATA TCACCAAACA GATAACAATC AAAAAGGGTA	3840
TTGATGGAGA GGAATCTAGT TCAGATGCGA CAAGTAATGC CAACCTAACT ATTTAAACCA	3900
AAGAATTGAA ATTGACAGAA GACCTAAGTA TTTCAGGTTT CAATAAGCA GAGATTACAG	3960
CCAAAGATGG TAGAGATTTA ACTATTGGCA ACAGTAATGA CGGTAACAGC GGTGCCGAAG	4020
CCAAAACAGT AACTTTTAAAC AATGTTAAAG ATTCAAAAAT CTCTGCTGAC GGTCACAATG	4080
TGACACTAAA TAGCAAAGTG AAAACATCTA GCAGCAATGG CGGACGTGAA AGCAATAGCG	4140

ACAACGATAC CGGCTTAACT ATTACTGCAA AAAATGTAGA AGTAAACAAA GATATTACTT	4200
CTCTCAAAAC AGTAAATATC ACCGCGTCGG AAAAGGTTAC CACCACAGCA GGCTCGACCA	4260
TTAACGCAAC AAATGGCAAA GCAAGTATTA CAACCAAAAC AGGTGATATC AGCGGTACGA	4320
TTTCCGGTAA CACGGTAAGT GTTAGCGCGA CTGGTGATTT AACCACTAAA TCCGGCTCAA	4380
AAATTGAAGC GAAATCGGGT GAGGCTAATG TAACAAGTGC AACAGGTACA ATTGGCGGTA	4440
CAATTTCCGG TAATACGGTA AATGTTACGG CAAACGCTGG CGATTTAACA GTTGGAATG	4500
GCGCAGAAAT TAATGCGACA GAAGGAGCTG CAACCTTAAC CGCAACAGGG AATACCTTGA	4560
CTACTGAAGC CGGTTCTAGC ATCACTTCAA CTAAGGGTCA GGTAGACCTC TTGGCTCAGA	4620
ATGGTAGCAT CGCAGGAAGC ATTAATGCTG CTAATGTGAC ATTAAATACT ACAGGCACCT	4680
TAACCACCGT GGCAGGCTCG GATATTAAAG CAACCAGCGG CACCTTGGTT ATTAACGCAA	4740
AAGATGCTAA GCTAAATGGT GATGCATCAG GTGATAGTAC AGAAGTGAAT GCAGTCAACG	4800
ACTGGGGATT TGGTAGTGTG ACTGCGGCAA CCTCAAGCAG TGTGAATATC ACTGGGGATT	4860
TAAACACAGT AAATGGGTTA AATATCATTT CGAAAGATGG TAGAAACACT GTGCGCTTAA	4920
GAGGCAAGGA AATTGAGGTG AAATATATCC AGCCAGGTGT AGCAAGTGTA GAAGAAGTAA	4980
TTGAAGCGAA ACGCGTCCTT GAAAAAGTAA AAGATTTATC TGATGAAGAA AGAGAAACAT	5040
TAGCTAAACT TGGTGTAAGT GCTGTACGTT TTGTTGAGCC AAATAATACA ATTACAGTCA	5100
ATACACAAAA TGAATTTACA ACCAGACCGT CAAGTCAAGT GATAATTTCT GAAGGTAAGG	5160
CGTGTTCCTC AAGTGGTAAT GCGCACGAG TATGTACCAA TGTGCTGAC GATGGACAGC	5220
CGTAGTCAGT AATTGACAAG GTAGATTTC A TCCTGCAATG AAGTCATTTT ATTTTCGTAT	5280
TATTTACTGT GTGGGTAAAA GTTCAGTACG GGCTTTACCC ATCTTGTA AAATTACGGA	5340
GAATACAATA AAGTATTTTT AACAGGTAT TATTATGAAA AATATAAAAA GCAGATTAAA	5400
ACTCAGTGCA ATATCAGTAT TGCTTGGCCT GGCTTCTTCA TCATTGTATG CAGAAGAAGC	5460
GTTTTTAGTA AAAGGCTTTC AGTTATCTGG TGCCTTGAA ACTTTAAGTG AAGACGCCCA	5520
ACTGTCTGTA GCAAAATCTT TATCTAAATA CCAAGGCTCG CAACTTTAA CAAACCTAAA	5580
AACAGCACAG CTTGAATTAC AGGCTGTGCT AGATAAGATT GAGCCAAATA AATTTGATGT	5640
GATATTGCCG CAACAAACCA TTACGGATGG CAATATCATG TTTGAGCTAG TCTCGAAATC	5700
AGCCGCAGAA AGCCAAGTTT TTTATAAGGC GAGCCAGGGT TATAGTGAAG AAAATATCGC	5760
TCGTAGCCTG CCATCTTTGA AACAAGGAAA AGTGTATGAA GATGGTCGTC AGTGGTTCGA	5820
TTTGCGTGAA TTTAATATGG CAAAAGAAAA CCCGCTTAAG GTTACCCGTG TACATTACGA	5880
ACTAAACCCT AAAAACAAAA CCTCTAATTT GATAATTGCG GGCTTCTCGC CTTTTGGTAA	5940
AACGCGTAGC TTTATTTCTT ATGATAATTT CGGCGCGAGA GAGTTTAACT ACCAACGTGT	6000
AAGCTTGGGT TTTGTTAATG CCAATTTAAC TGGTCATGAT GATGTGTAA TTATACCAGT	6060
ATGAGTTATG CTGATTCTAA TGATATCGAC GGCTTACCAA GTGCGATTAA TCGTAAATTA	6120
TCAAAAGGTC AATCTATCTC TGCGAATCTG AAATGGAGTT ATTATCTCCC AACATTTAAC	6180

CTTGGCATGG AAGACCAATT TAAAATTAAT TTAGGCTACA ACTACCGCCA TATTAATCAA	6240
ACCTCCGCGT TAAATCGCTT GGGTGAAACG AAGAAAAAAT TTGCAGTATC AGGCGTAAGT	6300
GCAGGCATTG ATGGACATAT CCAATTTACC CCTAAAACAA TCTTTAATAT TGATTTAACT	6360
CATCATTATT ACGCGAGTAA ATTACCAGGC TCTTTTGGA TGGAGCGCAT TGGCGAAACA	6420
TTTAATCGCA GCTATCACAT TAGCACAGCC AGTTTAGGGT TGAGTCAAGA GTTTGCTCAA	6480
GGTTGGCATT TTAGCAGTCA ATTATCAGGT CAATTTACTC TACAAGATAT TAGCAGTATA	6540
GATTTATTCT CTGTAACAGG TACTTATGGC GTCAGAGGCT TTAAATACGG CGGTGCAAGT	6600
GGTGAGCGCG GTCTTGATG GCGTAATGAA TTAAGTATGC CAAAATACAC CCGCTTCCAA	6660
ATCAGCCCTT ATGCGTTTTA TGATGCAGGT CAGTTCCGTT ATAATAGCGA AAATGCTAAA	6720
ACTTACGGCG AAGATATGCA CACGGTATCC TCTGCGGGT TAGGCATTAA AACCTCTCCT	6780
ACACAAAAC TAAACCTAGA TGCTTTTGTT GCTCGTCGCT TTGCAAAATGC CAATAGTGAC	6840
AATTTGAATG GCAACAAAAA ACGCACAAGC TCACCTACAA CCTTCTGGGG GAGATTAACA	6900
TTCAGTTTCT AACCTGAAA TTTAATCAAC TGTAAGCGT TCCGCCTACC AGTTTATAAC	6960
TATATGCTTT ACCCGCCAAT TTACAGTCTA TAGGCAACCC TGTTTTTACC CTTATATATC	7020
AAATAACAA GCTAAGCTGA GCTAAGCAA CCAAGCAAC TCAAGCAAGC CAAGTAATAC	7080
TAAAAAACA ATTTATATGA TAACTAAAG TATACTCCAT GCCATGGCGA TACAAGGGAT	7140
TTAATAATAT GACAAAAGAA AATTTGCAA ACGCTCCTCA AGATGCGACC GCTTTACTTG	7200
CGGAATTAAG CAACAATCAA ACTCCCCTGC GAATATTTAA ACAACCACGC AAGCCCAGCC	7260
TATTACGCTT GGAACAACAT ATCGCAAAAA AAGATTATGA GTTTGCTTGT CGTGAATTAA	7320
TGGTGATTCT GGAAAAAATG GACGCTAATT TTGGAGGCGT TCACGATATT GAATTTGACG	7380
CACCCGCTCA GCTGGCATAT CTACCCGAAA AATTACTAAT TTATTTTGCC ACTCGTCTCG	7440
CTAATGCAAT TACAACACTC TTTTCCGACC CCGAATTGGC AATTTCTGAA GAAGGGGCGT	7500
TAAAGATGAT TAGCCTGCAA CGCTGGTTGA CGCTGATTTT TGCCTCTTCC CCCTACGTTA	7560
ACGCAGACCA TATTCTCAAT AAATATAATA TCAACCCAGA TTCCGAAGGT GGCTTTCATT	7620
TAGCAACAGA CAACTCTTCT ATTGCTAAAT TCTGTATTTT TTAATTACCC GAATCCAATG	7680
TCAATATGAG TTTAGATGCG TTATGGGCAG GGAATCAACA ACTTTGTGCT TCATTGTGTT	7740
TTGCGTTGCA GTCTTCACGT TTTATTGGTA CCGCATCTGC GTTTCATAAA AGAGCGGTGG	7800
TTTTACAGTG GTTTCCTAAA AAACGCGCG AAATTGCTAA TTTAGATGAA TTGCCTGCAA	7860
ATATCCTTCA TGATGTATAT ATGCACTGCA GTTATGATTT AGCAAAAAAC AAGCACGATG	7920
TTAAGCGTCC ATTAAACGAA CTTGTCCGCA AGCATATCCT CACGCAAGGA TGGCAAGACC	7980
GCTACCTTTA CACCTTAGGT AAAAAGGACG GCAAACCTGT GATGATGGTA CTGCTTGAAC	8040
ATTTTAATTC GGGACATTCT ATTTATCGTA CACATTCAAC TTCAATGATT GCTGCTCGAG	8100
AAAAATTCTA TTTAGTCGGC TTAGGCCATG AGGGCGTTGA TAAAATAGGT CGAGAAGTGT	8160
TTGACGAGTT CTTTGAAATC AGTAGCAATA ATATAATGGA GAGACTGTTT TTTATCCGTA	8220



87

AACAGTGCGA AACTTTCCAA CCCGCAGTGT TCTATATGCC AAGCATTGGC ATGGATATTA	8280
CCACGATTTT TGTGAGCAAC ACTCGGCTTG CCCCTATTCA AGCTGTAGCC CTGGGTCATC	8340
CTGCCACTAC GCATTCTGAA TTTATTGATT ATGTCATCGT AGAAGATGAT TATGTGGGCA	8400
GTGAAGATTG TTTCAGCGAA ACCCTTTTAC GCTTACCCAA AGATGCCCTA CCTTATGTAC	8460
CTTCTGCACT CGCCCCACAA AAAGTGGATT ATGTACTCAG GGAAAACCTT GAAGTAGTCA	8520
ATATCGGTAT TGCCGCTACC ACAATGAAAT TAAACCTGA ATTTTGTCTA ACATTGCAAG	8580
AAATCAGAGA TAAAGCTAAA GTCAAAATAC ATTTTCATTT CGCACTTGGA CAATCAACAG	8640
GCTTGACACA CCCTTATGTC AAATGGTTTA TCGAAAGCTA TTTAGGTGAC GATGCCACTG	8700
CACATCCCCA CGCACCTTAT CACGATTATC TGGCAATATT GCGTGATTGC GATATGCTAC	8760
TAAATCCGTT TCCTTTCGGT AATACTAACG GCATAATTGA TATGGTTACA TTAGGTTTAG	8820
TTGGTGTATG CAAAACGGGG GATGAAGTAC ATGAACATAT TGATGAAGGT CTGTTTAAAC	8880
GCTTAGGACT ACCAGAATGG CTGATAGCCG ACACACGAGA AACATATATT GAATGTGCTT	8940
TGCGTCTAGC AGAAAACCAT CAAGAACGCC TTGAACTCCG TCGTTACATC ATAGAAAACA	9000
ACGGCTTACA AAAGCTTTTT ACAGGCGACC CTCGTCCATT GGGCAAAATA CTGCTTAAGA	9060
AAACAAATGA ATGGAAGCGG AAGCACTTGA GTAAAAAATA ACGGTTTTTT AAAGTAAAG	9120
TGCGGTAAAT TTTCAAAGCG TTTTAAAAAC CTCTCAAAAA TCAACCGCAC TTTTATCTTT	9180
ATAACGATCC CGCACGCTGA CAGTTTATCA GCCTCCCGCC ATAAACTCC GCCTTTCATG	9240
GCGGAGATTT TAGCCAAAAC TGGCAGAAAT TAAAGGCTAA AATCACCAAA TTGCACCACA	9300
AAATCACCAA TACCACAAA AAA	9323

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4794 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGAACAAGA TATATCGTCT CAAATTCAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT	60
GAATTGACAC GGGGTTGTGA CCATTCCACA GAAAAAGGCA GTGAAAAACC TGTTTCGTACG	120
AAAGTACGCC ACTTGGCGTT AAAGCCACTT TCCGCTATAT TGCTATCTTT GGGCATGGCA	180
TCCATTCGCG AATCTGTTTT AGCGAGCGGT TTACAGGGAA TGAGCGTCGT ACACGGTACA	240
GCAACCATGC AAGTAGACGG CAATAAAACC ACTATCCGTA ATAGCGTCAA TGCTATCATC	300
AATTGGAAAC AATTTAACAT TGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAGCAGC	360
AACTCTGCCG TTTTCAACCG TGTTACATCT GACCAAATCT CCCAATTAAA AGGGATTTTA	420

GATTCTAACG GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA	480
ATTATTAACA CTAATGGCTT TACTGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG	540
GCGCGTAATT TCACCCTTGA GCAAACCAAG GATAAAGCAC TCGCTGAAAT CGTGAATCAC	600
GGTTTAATTA CCGTTGGTAA AGACGGTAGC GTAAACCTTA TTGGTGGCAA AGTGAAAAAC	660
GAGGGCGTGA TTAGCGTAAA TGGCGGTAGT ATTTCTTTAC TTGCAGGGCA AAAAATCACC	720
ATCAGCGATA TAATAAATCC AACCATCACT TACAGCATTG CTGCACCTGA AAACGAAGCG	780
ATCAATCTGG GCGATATTTT TGCCAAAGGT GGTAACATTA ATGTCCGCGC TGCCACTATT	840
CGCAATAAAG GTAAACTTTC TGCCGACTCT GTAAGCAAAG ATAAAAGTGG TAACATTGTT	900
CTCTCTGCCA AAGAAGGTGA AGCGGAAATT GCGGGTGTA TTTCCGCTCA AAATCAGCAA	960
GCCAAAGGTG GTAAGTTGAT GATTACAGGC GATAAAGTTA CATTGAAAAC GGGTGCAGTT	1020
ATCGACCTTT CGGGTAAAGA AGGGGGAGAA ACTTATCTTG GCGGTGACGA GCGTGGCGAA	1080
GGTAAAAACG GCATTCAATT AGCAAAGAAA ACCACTTTAG AAAAAGGCTC AACAATTAAT	1140
GTGTCAGGTA AAGAAAAAGG TGGGCGCGCT ATTGTATGGG GCGATATTGC GTTAATTGAC	1200
GGCAATATTA ATGCCCAAGG TAAAGATATC GCTAAACTG GTGGTTTTGT GGAGACGTCG	1260
GGGCATTACT TATCCATTGA TGATAACGCA ATTGTAAAA CAAAAGAATG GCTACTAGAC	1320
CCAGAGAATG TGAATATTGA AGCTCCTTCC GCTTCTCGCG TCGAGCTGGG TGCCGATAGG	1380
AATTCCTACT CGGCAGAGGT GATAAAGTG ACCCTAAAA AAAATAACAC CTCCTTGACA	1440
ACACTAACCA ATACAACCAT TTCAAATCTT CTGAAAAGTG CCCACGTGGT GAACATAACG	1500
GCAAGGAGAA AACTTACCGT TAATAGCTCT ATCAGTATAG AAAGAGGCTC CCACTTAATT	1560
CTCCACAGTG AAGGTCAGGG CGGTCAAGGT GTTCAGATTG ATAAAGATAT TACTTCTGAA	1620
GGCGGAAATT TAACCATTTA TTCTGGCGGA TGGGTTGATG TTCATAAAAA TATTACGCTT	1680
GGTAGCGGCT TTTTAAACAT CACAATAAA GAAGGAGATA TCGCCTTCGA AGACAAGTCT	1740
GGACGGAACA ACCTAACCAT TACAGCCCAA GGGACCATCA CCTCAGGTAA TAGTAACGGC	1800
TTTAGATTTA ACAACGTCTC TCTAAACAGC CTTGGCGGAA AGCTGAGCTT TACTGACAGC	1860
AGAGAGGACA GAGGTAGAAG AACTAAGGGT AATATCTCAA ACAAATTTGA CGGAACGTTA	1920
AACATTTCCG GAACTGTAGA TATCTCAATG AAAGCACCCA AAGTCAGCTG GTTTTACAGA	1980
GACAAAGGAC GCACCTACTG GAACGTAACC ACTTTAAATG TTACCTCGGG TAGTAAATTT	2040
AACCTCTCCA TTGACAGCAC AGGAAGTGGC TCAACAGGTC CAAGCATACG CAATGCAGAA	2100
TTAAATGGCA TAACATTTAA TAAAGCCACT TTTAATATCG CACAAGGCTC AACAGCTAAC	2160
TTTAGCATCA AGGCATCAAT AATGCCCTTT AAGAGTAACG CTAACCTACG ATTATTTAAT	2220
GAAGATATTT CAGTCTCAGG GGGGGGTAGC CTTAATTTCA AACTTAACGC CTCATCTAGC	2280
AACATACAAA CCCCTGGCGT AATTATAAAA TCTCAAACT TTAATGTCTC AGGAGGGTCA	2340
ACTTTAAATC TCAAGGCTGA AGGTCAACA GAAACCGCTT TTTCAATAGA AAATGATTTA	2400
AACTTAAACG CCACCGGTGG CAATATAACA ATCAGACAAG TCGAGGGTAC CGATTACGCG	2460

GTCAACAAAG GTGTCGCAGC CAAAAAAAC ATAACCTTTA AAGGGGGTAA TATCACCTTC	2520
GGCTCTCAAA AAGCCACAAC AGAAATCAAA GGCAATGTTA CCATCAATAA AAACACTAAC	2580
GCTACTCTTT GTGGTGCGAA TTTTGCCGAA AACAAATCGC CTTTAAATAT AGCAGGAAAT	2640
GTTATTAATA ATGGCAACCT TACCACTGCC GGCTCCATTA TCAATATAGC CGGAAATCTT	2700
ACTGTTTCAA AAGGCGCTAA CCTTCAAGCT ATAACAAATT AACTTTTTAA TGTAGCCGGC	2760
TCATTGACA ACAATGGCGC TTCAAACATT TCCATTGCCA GAGGAGGGGC TAAATTTAAA	2820
GATATCAATA ACACCAGTAG CTTAAATATT ACCACCAACT CTGATACCAC TTACCGCACC	2880
ATTATAAAAG GCAATATATC CAACAAATCA GGTGATTGTA ATATTATTGA TAAAAAAGC	2940
GACGCTGAAA TCCAAATTGG CGGCAATATC TCACAAAAAG AAGGCAATCT CACAATTTCT	3000
TCTGATAAAG TAAATATTAC CAATCAGATA ACAATCAAAG CAGGCGTTGA AGGGGGGCGT	3060
TCTGATTCAA GTGAGGCAGA AAATGCTAAC CTAACATTC AAACCAAAGA GTTAAAATTG	3120
GCAGGAGACC TAAATATTTT AGGCTTTAAT AAAGCAGAAA TTACAGCTAA AAATGGCAGT	3180
GATTTAACTA TTGGCAATGC TAGCGGTGGT AATGCTGATG CTAAAAAAGT GACTTTTGAC	3240
AAGGTTAAAG ATTCAAAAAT CTCGACTGAC GGTCACAATG TAACACTAAA TAGCGAAGTG	3300
AAAACGTCTA ATGGTAGTAG CAATGCTGGT AATGATAACA GCACCGGTTT AACCATTTC	3360
GCAAAAGATG TAACGGTAAA CAATAACGTT ACCTCCCACA AGACAATAAA TATCTCTGCC	3420
GCAGCAGGAA ATGTAACAAC CAAAGAAGGC ACAACTATCA ATGCAACCAC AGGCAGCGTG	3480
GAAGTAACTG CTCAAAATGG TACAATTAAA GGCAACATTA CCTCGCAAAA TGTAACAGTG	3540
ACAGCAACAG AAAATCTTGT TACCACAGAG AATGCTGTCA TTAATGCAAC CAGCGGCACA	3600
GTAAACATTA GTACAAAAAC AGGGGATATT AAAGGTGGAA TTGAATCAAC TTCCGGTAAT	3660
GTAAATATTA CAGCGAGCGG CAATACACTT AAGGTAAGTA ATATCACTGG TCAAGATGTA	3720
ACAGTAACAG CGGATGCAGG AGCCTTGACA ACTACAGCAG GCTCAACCAT TAGTGCGACA	3780
ACAGGCAATG CAAATATTAC AACCAAAACA GGTGATATCA ACGGTAAAGT TGAATCCAGC	3840
TCCGGCTCTG TAACACTTGT TGCAACTGGA GCAACTCTTG CTGTAGGTAA TATTCAGGT	3900
AACACTGTTA CTATTACTGC SGATAGCGGT AAATTAACCT CCACAGTAGG TTCTACAATT	3960
AATGGGACTA ATAGTGTAAC CACCTCAAGC CAATCAGGCG ATATTGAAGG TACAATTTCT	4020
GGTAATACAG TAAATGTTAC AGCAAGCACT GGTGATTTAA CTATTGGAAA TAGTGCAAAA	4080
GTTGAAGCGA AAAATGGAGC TGCAACCTTA ACTGCTGAAT CAGGCAAATT AACCACCCAA	4140
ACAGGCTCTA GCATTACCTC AAGCAATGGT CAGACAACCT TTACAGCCAA GGATAGCAGT	4200
ATCGCAGGAA ACATTAATGC TGCTAATGTG ACGTTAAATA CCACAGGCAC TTAACTACT	4260
ACAGGGGATT CAAAGATTAA CGCAACCAGT GGTACCTTAA CAATCAATGC AAAAGATGCC	4320
AAATTAGATG GTGCTGCATC AGGTGACCGC ACAGTAGTAA ATGCAACTAA CGCAAGTGCC	4380
TCTGGTAACG TGAATGCGAA AACCTCAAGC AGCGTGAATA TCACCGGGGA TTAAACACA	4440
ATAAATGGGT TAAATATCAT TTCGGAAAAT GGTAGAAACA CTGTGCGCTT AAGAGGCAAG	4500

GAAATTGATG TGAAATATAT CCAACCAGGT GTAGCAAGCG TAGAAGAGGT AATTGAAGCG	4560
AAACGCGTCC TTGAGAAGGT AAAAGATTTA TCTGATGAAG AAAGAGAAAC ACTAGCCAAA	4620
CTTGGTGTAA GTGCTGTACG TTTCGTTGAG CCAAATAATG CCATTACGGT TAATACACAA	4680
AACGAGTTTA CAACCAAACC ATCAAGTCAA GTGACAATTT CTGAAGGTAA GCGGTGTTTC	4740
TCAAGTGGTA ATGGCGCACG AGTATGTACC AATGTTGCTG ACGATGGACA GCAG	4794

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4803 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGAACAAGA TATATCGTCT CAAATTCAGC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT	60
GAATTGACAC GGGGTGTGA CCATTCCACA GAAAAAGGCA GTGAAAAACC TGTCGTACG	120
AAAGTACGCC ACTTGGCGTT AAAGCCACTT TCCGCTATAT TGCTATCTTT GGGCATGGCA	180
TCCATTCCGC AATCTGTTTT AGCGAGCGGT TTACAGGGAA TGAGCGTCGT ACACGGTACA	240
GCAACCATGC AAGTAGACGG CAATAAAACC ACTATCCGTA ATAGCGTCAA TGCTATCATC	300
AATTGAAAC AATTTAACAT TGACCAAAT GAAATGGTGC AGTTTTTACA AGAAAGCAGC	360
AACTCTGCCG TTTTCAACCG TGTTACATCT GACCAAATCT CCCAATTAAA AGGGATTTTA	420
GATTCTAACG GACAAGTCTT TTTAATCAAC CCAATGGTA TCACAATAGG TAAAGACGCA	480
ATTATTAACA CTAATGGCTT TACTGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG	540
GCGCGTAATT TCACCCTTGA GCAAACCAAG GATAAAGCAC TCGCTGAAAT CGTGAATCAC	600
GGTTTAATTA CCGTTGGTAA AGACGGTAGC GTAAACCTTA TTGGTGGCAA AGTGAAAAAC	660
GAGGGCGTGA TTAGCGTAAA TGGCGGTAGT ATTTCTTTAC TTGCAGGGCA AAAAATCACC	720
ATCAGCGATA TAATAAATCC AACCATCACT TACAGCATTG CTGCACCTGA AAACGAAGCG	780
ATCAATCTGG GCGATATTTT TGCCAAAGGT GGTAACATTA ATGTCCGCGC TGCCACTATT	840
CGCAATAAAG GTAACTTTC TGCCGACTCT GTAAGCAAAG ATAAAGTGG TAACATTGTT	900
CTCTCTGCCA AAGAAGGTGA AGCGGAAATT GGCGGTGTAA TTTCCGCTCA AAATCAGCAA	960
GCCAAAGGTG GTAAGTTGAT GATTACAGGT GATAAAGTCA CATTAAAAAC AGGTGCAGTT	1020
ATCGACCTTT CAGGTAAAGA AGGGGAGAG ACTTATCTTG GCGGTGATGA GCGTGGCGAA	1080
GGTAAAAATG GTATTCAATT AGCGAAGAAA ACCTCTTTAG AAAAAGGCTC GACAATTAAT	1140
GTATCAGGCA AAGAAAAAGG CGGGCGCGCT ATTGTATGGG GCGATATTGC ATTAATTAAT	1200
GGTAACATTA ATGCTCAAGG TAGCGATATT GCTAAACTG GCGGCTTTGT GGAAACATCA	1260

GGACATGACT TATCCATTGG TGATGATGTG ATTGTTGACG CTAAAGAGTG GTTATTAGAC	1320
CCAGATGATG TGTCCATTGA AACTCTTACA TCTGGACGCA ATAATACCGG CGAAAACCAA	1380
GGATATACAA CAGGAGATGG GACTAAAGAG TCACCTAAAG GTAATAGTAT TTCTAAACCT	1440
ACATTAACAA ACTCAACTCT TGAGCAAATC CTAAGAAGAG GTTCTTATGT TAATATCACT	1500
GCTAATAATA GAATTTATGT TAATAGCTCC ATCAACTTAT CTAATGGCAG TTTAACACTT	1560
CACACTAAAC GAGATGGAGT TAAAATTAAC GGTGATATTA CCTCAAACGA AAATGGTAAT	1620
TTAACCATTA AAGCAGGCTC TTGGGTTGAT GTTCATAAAA ACATCACGCT TGGTACGGGT	1680
TTTTTGAATA TTGTGCGTGG GGATTCTGTA GCTTTTGAGA GAGAGGGCGA TAAAGCACGT	1740
AACGCAACAG ATGCTCAAAT TACCGCACAA GGGACGATAA CCGTCAATAA AGATGATAAA	1800
CAATTTAGAT TCAATAATGT ATCTATTAAC GGGACGGGCA AGGGTTTAAA GTTTATTGCA	1860
AATCAAAATA ATTTCACTCA TAAATTTGAT GGCGAAATTA ACATATCTGG AATAGTAACA	1920
ATTAACCAAA CCACGAAAAA AGATGTTAAA TACTGGAATG CATCAAAAGA CTCTTACTGG	1980
AATGTTTCTT CTCTTACTTT GAATACGGTG CAAAAATTTA CCTTTATAAA ATTCGTTGAT	2040
AGCGGCTCAA ATTCCCAAGA TTTGAGGTCA TCACGTAGAA GTTTTGCAGG CGTACATTTT	2100
AACGGCATCG GAGGCAAAAC AAATTCAAC ATCGGAGCTA ACGCAAAAGC CTTATTTAAA	2160
TTAAAACCAA ACGCCGCTAC AGACCCAAAA AAAGAATTAC CTATTACTTT TAACGCCAAC	2220
ATTACAGCTA CCGGTAACAG TGATAGCTCT GTGATGTTTG ACATACACGC CAATCTTACC	2280
TCTAGAGCTG CCGGCATAAA CATGGATTCA ATTAACATTA CCGGCGGGCT TGACTTTTCC	2340
ATAACATCCC ATAATCGCAA TAGTAATGCT TTTGAAATCA AAAAAGACTT AACTATAAAT	2400
GCAACTGGCT CGAATTTTAG TCTTAAGCAA ACGAAAGATT CTTTTTATAA TGAATACAGC	2460
AAACACGCCA TTAATCAAG TCATAATCTA ACCATTCTTG GCGGCAATGT CACTCTAGGT	2520
GGGGAAAATT CAAGCAGTAG CATTACGGGC AATATCAATA TCACCAATAA AGCAAATGTT	2580
ACATTACAAG CTGACACCAG CAACAGCAAC ACAGGCTTGA AGAAAAGAAC TCTAACTCTT	2640
GGCAATATAT CTGTTGAGGG GAATTTAAGC CTAAGTGGTG CAAATGCAAA CATTGTCGGC	2700
AATCTTTCTA TTGCAGAAGA TTCCACATTT AAAGGAGAAG CCAGTGACAA CCTAAACATC	2760
ACCGGCACCT TTACCAACAA CGGTACCGCC AACATTAATA TAAAACAAGG AGTGGTAAAA	2820
CTCCAAGGCG ATATTATCAA TAAAGGTGGT TTAAATATCA CTACTAACGC CTCAGGCACT	2880
CAAAAAACCA TTATTAACGG AAATATAACT AACGAAAAAG GCGACTTAAA CATCAAGAAT	2940
ATTAAAGCCG ACGCCGAAAT CCAATTGGC GGCAATATCT CACAAAAAGA AGGCAATCTC	3000
ACAATTTCTT CTGATAAAGT AAATATTACC AATCAGATAA CAATCAAAGC AGGCGTTGAA	3060
GGGGGGCGTT CTGATTCAAG TGAGGCAGAA AATGCTAACC TAACTATTCA AACCAAAGAG	3120
TTAAAATTGG CAGGAGACCT AAATATTTCA GGCTTTAATA AAGCAGAAAT TACAGCTAAA	3180
AATGGCAGTG ATTTAACTAT TGCAATGCT AGCGGTGGTA ATGCTGATGC TAAAAAAGTG	3240
ACTTTTGACA AGGTAAAGA TTCAAAAATC TCGACTGACG GTCACAATGT AACACTAAAT	3300

92

AGCGAAGTGA AAACGTCTAA TGGTAGTAGC AATGCTGGTA ATGATAACAG CACCGGTTTA 3360  
 ACCATTTCCG CAAAAGATGT AACGGTAAAC AATAAGTTA CCTCCCACAA GACAATAAAT 3420  
 ATCTCTGCCG CAGCAGGAAA TGTAACAACC AAAGAAGGCA CAACTATCAA TGCAACCACA 3480  
 GGCAGCGTGG AAGTAACTGC TCAAAATGGT ACAATTAAG GCAACATTAC CTCGCAAAAT 3540  
 GTAACAGTGA CAGCAACAGA AAATCTTGTT ACCACAGAGA ATGCTGTCAT TAATGCAACC 3600  
 AGCGGCACAG TAAACATTAG TACAAAAACA GGGGATATTA AAGGTGGAAT TGAATCAACT 3660  
 TCCGTAATG TAAATATTAC AGCGAGCGGC AATACACTTA AGGTAAGTAA TATCACTGGT 3720  
 CAAGATGTAA CAGTAACAGC GGATGCAGGA GCCTTGACAA CTACAGCAGG CTCAACCATT 3780  
 AGTGCGACAA CAGGCAATGC AAATATTACA ACCAAAACAG GTGATATCAA CGGTAAAGTT 3840  
 GAATCCAGCT CCGGCTCTGT AACACTTGTT GCAACTGGAG CAACTCTTGC TGTAGGTAAT 3900  
 ATTCAGGTA AACTGTTAC TATTACTGCG GATAGCGGTA AATTAACCTC CACAGTAGGT 3960  
 TCTACAATTA ATGGGACTAA TAGTGTAACC ACCTCAAGCC AATCAGGCGA TATTGAAGGT 4020  
 ACAATTTCTG GTAATACAGT AAATGTTACA GCAAGCACTG GTGATTAAAC TATTGGAAAT 4080  
 AGTGCAAAAG TTGAAGCGAA AAATGGAGCT GCAACCTTAA CTGCTGAATC AGGCAATTA 4140  
 ACCACCCAAA CAGGCTCTAG CATTACCTCA AGCAATGGTC AGACAACCTCT TACAGCCAAG 4200  
 GATAGCAGTA TCGCAGGAAA CATTAATGCT GCTAATGTGA CGTTAAATAC CACAGGCACT 4260  
 TTAACACTA CAGGGGATTC AAAGATTAAC GCAACCAGTG GTACCTTAAC AATCAATGCA 4320  
 AAAGATGCCA AATTAGATGG TGCTGCATCA GGTGACCGCA CAGTAGTAAA TGCAACTAAC 4380  
 GCAAGTGGCT CTGGTAACGT GACTGCGAAA ACCTCAAGCA GCGTGAATAT CACCGGGGAT 4440  
 TTAAACACAA TAAATGGGTT AAATATCATT TCGGAAAATG GTAGAAACAC TGTGCGCTTA 4500  
 AGAGGCAAGG AAATTGATGT GAAATATATC CAACCAGGTG TAGCAAGCGT AGAAGAGGTA 4560  
 ATTGAAGCGA AACGCGTCCT TGAGAAGGTA AAAGATTTAT CTGATGAAGA AAGAGAAACA 4620  
 CTAGCCAAAC TTGGTGTAAG TGCTGTACGT TTCGTTGAGC CAAATAATGC CATTACGGTT 4680  
 AATACACAAA ACGAGTTTAC AACCAAACCA TCAAGTCAAG TGACAATTTT TGAAGGTAAG 4740  
 GCGTGTTTCT CAAGTGGTAA TGGCGCACGA GTATGTACCA ATGTTGCTGA CGATGGACAG 4800  
 CAG 4803

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1599 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu  
 1 5 10 15  
 Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys  
 20 25 30  
 Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys  
 35 40 45  
 Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln  
 50 55 60  
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr  
 65 70 75 80  
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val  
 85 90 95  
 Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
 100 105 110  
 Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val  
 115 120 125  
 Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
 130 135 140  
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
 145 150 155 160  
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
 165 170 175  
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys  
 180 185 190  
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
 195 200 205  
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
 210 215 220  
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
 225 230 235 240  
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
 245 250 255  
 Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
 260 265 270  
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala  
 275 280 285  
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
 290 295 300  
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
 305 310 315 320  
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
 325 330 335

Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
 340 345 350  
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
 355 360 365  
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
 370 375 380  
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
 385 390 395 400  
 Gly Asn Ile Asn Ala Gln Gly Lys Asp Ile Ala Lys Thr Gly Gly Phe  
 405 410 415  
 Val Glu Thr Ser Gly His Tyr Leu Ser Ile Asp Asp Asn Ala Ile Val  
 420 425 430  
 Lys Thr Lys Glu Trp Leu Leu Asp Pro Glu Asn Val Thr Ile Glu Ala  
 435 440 445  
 Pro Ser Ala Ser Arg Val Glu Leu Gly Ala Asp Arg Asn Ser His Ser  
 450 455 460  
 Ala Glu Val Ile Lys Val Thr Leu Lys Lys Asn Asn Thr Ser Leu Thr  
 465 470 475 480  
 Thr Leu Thr Asn Thr Thr Ile Ser Asn Leu Leu Lys Ser Ala His Val  
 485 490 495  
 Val Asn Ile Thr Ala Arg Arg Lys Leu Thr Val Asn Ser Ser Ile Ser  
 500 505 510  
 Ile Glu Arg Gly Ser His Leu Ile Leu His Ser Glu Gly Gln Gly Gly  
 515 520 525  
 Gln Gly Val Gln Ile Asp Lys Asp Ile Thr Ser Glu Gly Gly Asn Leu  
 530 535 540  
 Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr Leu  
 545 550 555 560  
 Gly Ser Gly Phe Leu Asn Ile Thr Thr Lys Glu Gly Asp Ile Ala Phe  
 565 570 575  
 Glu Asp Lys Ser Gly Arg Asn Asn Leu Thr Ile Thr Ala Gln Gly Thr  
 580 585 590  
 Ile Thr Ser Gly Asn Ser Asn Gly Phe Arg Phe Asn Asn Val Ser Leu  
 595 600 605  
 Asn Ser Leu Gly Gly Lys Leu Ser Phe Thr Asp Ser Arg Glu Asp Arg  
 610 615 620  
 Gly Arg Arg Thr Lys Gly Asn Ile Ser Asn Lys Phe Asp Gly Thr Leu  
 625 630 635 640  
 Asn Ile Ser Gly Thr Val Asp Ile Ser Met Lys Ala Pro Lys Val Ser  
 645 650 655  
 Trp Phe Tyr Arg Asp Lys Gly Arg Thr Tyr Trp Asn Val Thr Thr Leu  
 660 665 670  
 Asn Val Thr Ser Gly Ser Lys Phe Asn Leu Ser Ile Asp Ser Thr Gly  
 675 680 685



95

Ser Gly Ser Thr Gly Pro Ser Ile Arg Asn Ala Glu Leu Asn Gly Ile  
 690 695 700  
 Thr Phe Asn Lys Ala Thr Phe Asn Ile Ala Gln Gly Ser Thr Ala Asn  
 705 710 715 720  
 Phe Ser Ile Lys Ala Ser Ile Met Pro Phe Lys Ser Asn Ala Asn Tyr  
 725 730 735  
 Ala Leu Phe Asn Glu Asp Ile Ser Val Ser Gly Gly Gly Ser Val Asn  
 740 745 750  
 Phe Lys Leu Asn Ala Ser Ser Ser Asn Ile Gln Thr Pro Gly Val Ile  
 755 760 765  
 Ile Lys Ser Gln Asn Phe Asn Val Ser Gly Gly Ser Thr Leu Asn Leu  
 770 775 780  
 Lys Ala Glu Gly Ser Thr Glu Thr Ala Phe Ser Ile Glu Asn Asp Leu  
 785 790 795 800  
 Asn Leu Asn Ala Thr Gly Gly Asn Ile Thr Ile Arg Gln Val Glu Gly  
 805 810 815  
 Thr Asp Ser Arg Val Asn Lys Gly Val Ala Ala Lys Lys Asn Ile Thr  
 820 825 830  
 Phe Lys Gly Gly Asn Ile Thr Phe Gly Ser Gln Lys Ala Thr Thr Glu  
 835 840 845  
 Ile Lys Gly Asn Val Thr Ile Asn Lys Asn Thr Asn Ala Thr Leu Arg  
 850 855 860  
 Gly Ala Asn Phe Ala Glu Asn Lys Ser Pro Leu Asn Ile Ala Gly Asn  
 865 870 875 880  
 Val Ile Asn Asn Gly Asn Leu Thr Thr Ala Gly Ser Ile Ile Asn Ile  
 885 890 895  
 Ala Gly Asn Leu Thr Val Ser Lys Gly Ala Asn Leu Gln Ala Ile Thr  
 900 905 910  
 Asn Tyr Thr Phe Asn Val Ala Gly Ser Phe Asp Asn Asn Gly Ala Ser  
 915 920 925  
 Asn Ile Ser Ile Ala Arg Gly Gly Ala Lys Phe Lys Asp Ile Asn Asn  
 930 935 940  
 Thr Ser Ser Leu Asn Ile Thr Thr Asn Ser Asp Thr Thr Tyr Arg Thr  
 945 950 955 960  
 Ile Ile Lys Gly Asn Ile Ser Asn Lys Ser Gly Asp Leu Asn Ile Ile  
 965 970 975  
 Asp Lys Lys Ser Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser Gln  
 980 985 990  
 Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr Asn  
 995 1000 1005  
 Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser Ser  
 1010 1015 1020  
 Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys Leu  
 1025 1030 1035 1040

Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala  
 1045 1050 1055  
 Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn Ala  
 1060 1065 1070  
 Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile Ser  
 1075 1080 1085  
 Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser Asn  
 1090 1095 1100  
 Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile Ser  
 1105 1110 1115 1120  
 Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr Ile  
 1125 1130 1135  
 Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr Thr  
 1140 1145 1150  
 Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly Thr  
 1155 1160 1165  
 Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr Glu  
 1170 1175 1180  
 Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly Thr  
 1185 1190 1195 1200  
 Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu Ser  
 1205 1210 1215  
 Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys Val  
 1220 1225 1230  
 Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly Ala  
 1235 1240 1245  
 Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn Ala  
 1250 1255 1260  
 Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser Ser  
 1265 1270 1275 1280  
 Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val Gly  
 1285 1290 1295  
 Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys Leu  
 1300 1305 1310  
 Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr Thr  
 1315 1320 1325  
 Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr Val  
 1330 1335 1340  
 Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala Lys  
 1345 1350 1355 1360  
 Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly Lys  
 1365 1370 1375  
 Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln Thr  
 1380 1385 1390

97

Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala Ala  
 1395 1400 1405  
 Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp Ser  
 1410 1415 1420  
 Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp Ala  
 1425 1430 1435 1440  
 Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala Thr  
 1445 1450 1455  
 Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser Val  
 1460 1465 1470  
 Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile Ser  
 1475 1480 1485  
 Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp Val  
 1490 1495 1500  
 Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu Ala  
 1505 1510 1515 1520  
 Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg Glu  
 1525 1530 1535  
 Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro Asn  
 1540 1545 1550  
 Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro Ser  
 1555 1560 1565  
 Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly Asn  
 1570 1575 1580  
 Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro  
 1585 1590 1595

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1600 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu  
 1 5 10 15  
 Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys  
 20 25 30  
 Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys  
 35 40 45  
 Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln  
 50 55 60  
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr  
 65 70 75 80

98

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val  
85 90 95

Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
100 105 110

Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val  
115 120 125

Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys  
180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
245 250 255

Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala  
275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
290 295 300

Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
305 310 315 320

Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
325 330 335

Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
340 345 350

Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
355 360 365

Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
370 375 380

Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
385 390 395 400

Gly Asn Ile Asn Ala Gln Gly Ser Asp Ile Ala Lys Thr Gly Gly Phe  
405 410 415

Val Glu Thr Ser Gly His Asp Leu Ser Ile Gly Asp Asp Val Ile Val  
420 425 430

Asp Ala Lys Glu Trp Leu Leu Asp Pro Asp Asp Val Ser Ile Glu Thr  
 435 440 445  
 Leu Thr Ser Gly Arg Asn Asn Thr Gly Glu Asn Gln Gly Tyr Thr Thr  
 450 455 460  
 Gly Asp Gly Thr Lys Glu Ser Pro Lys Gly Asn Ser Ile Ser Lys Pro  
 465 470 475 480  
 Thr Leu Thr Asn Ser Thr Leu Glu Gln Ile Leu Arg Arg Gly Ser Tyr  
 485 490 495  
 Val Asn Ile Thr Ala Asn Asn Arg Ile Tyr Val Asn Ser Ser Ile Asn  
 500 505 510  
 Leu Ser Asn Gly Ser Leu Thr Leu His Thr Lys Arg Asp Gly Val Lys  
 515 520 525  
 Ile Asn Gly Asp Ile Thr Ser Asn Glu Asn Gly Asn Leu Thr Ile Lys  
 530 535 540  
 Ala Gly Ser Trp Val Asp Val His Lys Asn Ile Thr Leu Gly Thr Gly  
 545 550 555 560  
 Phe Leu Asn Ile Val Ala Gly Asp Ser Val Ala Phe Glu Arg Glu Gly  
 565 570 575  
 Asp Lys Ala Arg Asn Ala Thr Asp Ala Gln Ile Thr Ala Gln Gly Thr  
 580 585 590  
 Ile Thr Val Asn Lys Asp Asp Lys Gln Phe Arg Phe Asn Asn Val Ser  
 595 600 605  
 Leu Asn Gly Thr Gly Lys Gly Leu Lys Phe Ile Ala Asn Gln Asn Asn  
 610 615 620  
 Phe Thr His Lys Phe Asp Gly Glu Ile Asn Ile Ser Gly Ile Val Thr  
 625 630 635 640  
 Ile Asn Gln Thr Thr Lys Lys Asp Val Lys Tyr Trp Asn Ala Ser Lys  
 645 650 655  
 Asp Ser Tyr Trp Asn Val Ser Ser Leu Thr Leu Asn Thr Val Gln Lys  
 660 665 670  
 Phe Thr Phe Ile Lys Phe Val Asp Ser Gly Ser Asn Gly Gln Asp Leu  
 675 680 685  
 Arg Ser Ser Arg Arg Ser Phe Ala Gly Val His Phe Asn Gly Ile Gly  
 690 695 700  
 Gly Lys Thr Asn Phe Asn Ile Gly Ala Asn Ala Lys Ala Leu Phe Lys  
 705 710 715 720  
 Leu Lys Pro Asn Ala Ala Thr Asp Pro Lys Lys Glu Leu Pro Ile Thr  
 725 730 735  
 Phe Asn Ala Asn Ile Thr Ala Thr Gly Asn Ser Asp Ser Ser Val Met  
 740 745 750  
 Phe Asp Ile His Ala Asn Leu Thr Ser Arg Ala Ala Gly Ile Asn Met  
 755 760 765  
 Asp Ser Ile Asn Ile Thr Gly Gly Leu Asp Phe Ser Ile Thr Ser His  
 770 775 780

100

Asn Arg Asn Ser Asn Ala Phe Glu Ile Lys Lys Asp Leu Thr Ile Asn  
 785 790 795 800  
 Ala Thr Gly Ser Asn Phe Ser Leu Lys Gln Thr Lys Asp Ser Phe Tyr  
 805 810 815  
 Asn Glu Tyr Ser Lys His Ala Ile Asn Ser Ser His Asn Leu Thr Ile  
 820 825 830  
 Leu Gly Gly Asn Val Thr Leu Gly Gly Glu Asn Ser Ser Ser Ile  
 835 840 845  
 Thr Gly Asn Ile Asn Ile Thr Asn Lys Ala Asn Val Thr Leu Gln Ala  
 850 855 860  
 Asp Thr Ser Asn Ser Asn Thr Gly Leu Lys Lys Arg Thr Leu Thr Leu  
 865 870 875 880  
 Gly Asn Ile Ser Val Glu Gly Asn Leu Ser Leu Thr Gly Ala Asn Ala  
 885 890 895  
 Asn Ile Val Gly Asn Leu Ser Ile Ala Glu Asp Ser Thr Phe Lys Gly  
 900 905 910  
 Glu Ala Ser Asp Asn Leu Asn Ile Thr Gly Thr Phe Thr Asn Asn Gly  
 915 920 925  
 Thr Ala Asn Ile Asn Ile Lys Gly Val Val Lys Leu Gly Asp Ile Asn  
 930 935 940  
 Asn Lys Gly Gly Leu Asn Ile Thr Thr Asn Ala Ser Gly Thr Gln Lys  
 945 950 955 960  
 Thr Ile Ile Asn Gly Asn Ile Thr Asn Glu Lys Gly Asp Leu Asn Ile  
 965 970 975  
 Lys Asn Ile Lys Ala Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser  
 980 985 990  
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr  
 995 1000 1005  
 Asn Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser  
 1010 1015 1020  
 Ser Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys  
 1025 1030 1035 1040  
 Leu Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr  
 1045 1050 1055  
 Ala Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn  
 1060 1065 1070  
 Ala Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile  
 1075 1080 1085  
 Ser Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser  
 1090 1095 1100  
 Asn Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile  
 1105 1110 1115 1120  
 Ser Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr  
 1125 1130 1135

101

Ile Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr  
 1140 1145 1150  
 Thr Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly  
 1155 1160 1165  
 Thr Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr  
 1170 1175 1180  
 Glu Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly  
 1185 1190 1195 1200  
 Thr Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu  
 1205 1210 1215  
 Ser Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys  
 1220 1225 1230  
 Val Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly  
 1235 1240 1245  
 Ala Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn  
 1250 1255 1260  
 Ala Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser  
 1265 1270 1275 1280  
 Ser Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val  
 1285 1290 1295  
 Gly Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys  
 1300 1305 1310  
 Leu Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr  
 1315 1320 1325  
 Thr Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr  
 1330 1335 1340  
 Val Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala  
 1345 1350 1355 1360  
 Lys Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly  
 1365 1370 1375  
 Lys Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln  
 1380 1385 1390  
 Thr Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala  
 1395 1400 1405  
 Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp  
 1410 1415 1420  
 Ser Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp  
 1425 1430 1435 1440  
 Ala Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala  
 1445 1450 1455  
 Thr Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser  
 1460 1465 1470  
 Val Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile  
 1475 1480 1485

102

Ser Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp  
 1490 1495 1500  
 Val Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu  
 1505 1510 1515 1520  
 Ala Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg  
 1525 1530 1535  
 Glu Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro  
 1540 1545 1550  
 Asn Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro  
 1555 1560 1565  
 Ser Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly  
 1570 1575 1580  
 Asn Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro  
 1585 1590 1595 1600

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Val Asp Glu Val Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp  
 1 5 10 15  
 Leu Ser Asp Glu Glu Arg Glu Ala Leu Ala Lys Leu Gly  
 20 25



## SEQUENCE LISTING

<110> Barenkamp, Stephen J.

<120> HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE  
HAEMOPHILUS

<130>

<140>

<141>

<150> 09/155,614

<151> 1998-09-30

<150> 08/617,697

<151> 1996-04-01

<150> PCT/US97/04707

<151> 1997-04-01

<160> 11

<170> PatentIn Ver. 2.1

<210> 1

<211> 5116

<212> DNA

<213> Haemophilus influenzae

<400> 1

```

acagcgttct ctttaacta gtacaaaccc acaataaaat atgacaaaca acaattacaa 60
cacctttttt gcagtctata tgcaaatatt ttaaaaaata gtataaatcc gccatataaa 120
atggtataat ctttcatctt tcattctttc tctttcatct ttcattcttc atctttcatc 180
tttcatcttt catctttcat ctttcatctt tcattctttc tctttcatct ttcattcttc 240
acatgccctg atgaaccgag ggaaggagg gaggggcaag aatgaagagg gagctgaacg 300
aacgcaaatg ataaagtaat ttaattgttc aactaacctt aggagaaaat atgaacaagc 360
tatatcgtct caaattcagc aaacgcctga atgctttggg tgctgtgtct gaattggcac 420
ggggttggtg ccattccaca gaaaaaggca gcgaaaaacc tgctcgcatg aaagtgcgtc 480
acttagcggt aaagccactt tccgctatgt tactatcttt aggtgtaaca tctattccac 540
aatctgtttt agcaagcggc ttacaaggaa tggatgtagt acacggcaca gccactatgc 600
aagtagatgg taataaaacc attatccgca acagtgttga cgatatcatt aattggaaac 660
aatttaacat cgaccaaact gaaatggtgc agtttttaca agaaaacaac aactccgccg 720
tattcaaccg tgttacatct aaccaaactc ccaatttaa agggatttta gattctaacg 780
gacaagtctt tttaatcaac ccaaattgta tcacaatagg taaagacgca attattaaca 840
ctaattggct tacggcttct acgctagaca tttctaacga aaacatcaag gcgcgtaatt 900
tcaccttcga gcaaaccaaa gataaagcgc tcgctgaaat tgtgaatcac ggtttaatta 960
ctgtcggtta agacggcagt gtaaacttta ttggtggcaa agtgaaaaac gaggggtgta 1020
ttagcgtaaa tgggtggcagc atttctttac tcgcagggca aaaaatcacc atcagcgata 1080
taataaacc aaccattact tacagcattg ccgcgcctga aaatgaagcg gtcaatctgg 1140
gcgatatttt tgccaaaggc ggtaacatta atgtccgtgc tgccactatt cgaaaccaag 1200
gtaaactttc tgctgattct gtaagcaaag ataaaagcgg caatattgtt ctttccgcca 1260
aagaggggtg agcggaaatt ggcgggtgta tttccgctca aaatcagcaa gctaaaggcg 1320
gcaagctgat gattacaggc gataaagtca cattaaaaac aggtgcagtt atcgaccttt 1380
caggtaaaga agggggagaa acttaccttg gcggtgacga gcgcggcgaa ggtaaaaagg 1440
gcattcaatt agcaaagaaa acctcttttag aaaaaggctc aaccatcaat gtatcaggca 1500
aagaaaaagg cggacgcgct attgtgtggg gcgatattgc gtttaattgac ggcaatatta 1560
acgctcaagg tagtggtgat atcgctaaaa ccggtggttt tgtggagacg tcggggcatg 1620
atatttcat caaagacaat gcaattgttg acgcaaaga gtggtgttta gaccgggata 1680
atgtatctat taatgcagaa acagcaggac gcagcaatac ttcagaagac gatgaatata 1740
cgggatccgg gaatagtgcc agcaccctaa aacgaaacaa agaaaagaca acattaacaa 1800

```

acacaactct	tgagagtata	ctaaaaaaag	gtacctttgt	taacatcact	gctaatacaac	1860
gcatctatgt	caatagctcc	attaatttat	ccaatggcag	cttaactctt	tggagtgagg	1920
gtcggagcgg	tggcggcggt	gagattaaca	acgatattac	caccggtgat	gataccagag	1980
gtgcaaactt	aacaatttac	tcaggcggct	gggttgatgt	tcataaaaaat	atctcactcg	2040
gggcgcaagg	taacataaac	attacagcta	aacaagatat	cgcttttgag	aaaggaagca	2100
accaagtcat	tacagggtcaa	gggactatta	cctcaggcaa	tcaaaaagggt	tttagattta	2160
ataatgtctc	tctaaacggc	actggcagcg	gactgcaatt	caccactaaa	agaaccaata	2220
aatacgctat	cacaaataaaa	tttgaaggga	ctttaaatat	ttcagggaaa	gtgaacatct	2280
caatggtttt	acctaataaat	gaaagtggat	atgataaatt	caaaggacgc	acttactgga	2340
atttaacctc	cttaaatgtt	tccgagagt	gcgagtttaa	cctcactatt	gactccagag	2400
gaagcgatag	tgcaggcaca	cttaccacgc	cttataattt	aaacggtata	tcattcaaca	2460
aagacactac	ctttaatgtt	gaacgaaatg	caagagtcaa	ctttgacatc	aaggcaccac	2520
tagggataaa	taagtattct	agtttgaatt	acgcatcatt	taatggaaac	atttcagttt	2580
cgggaggggg	gagtgttgat	ttcacacttc	tcgcctcatc	ctctaacgtc	caaacccccg	2640
gtgtagttat	aaattctaaa	tacttttaatg	tttcaacagg	gtcaagttta	agatttataaa	2700
cttcaggctc	aacaaaaact	ggcttctcaa	tagagaaaga	tttaacttta	aatgccaccg	2760
gaggcaacat	aacacttttg	caagttgaag	gcaccgatgg	aatgattggg	aaaggcattg	2820
tagccaaaaa	aaacataacc	tttgaaggag	gtaacatcac	ctttggctcc	aggaaagccg	2880
taacagaaat	cgaaggcaat	gttactatca	ataacaacgc	taacgtcact	cttatcggtt	2940
cggattttga	caaccatcaa	aaacctttta	ctattaaaaa	agatgtcatc	attaatagcg	3000
gcaaccttac	cgctggaggc	aatattgtca	atatagccgg	aaatcttacc	gttgaaagta	3060
acgctaattt	caaagctatc	acaaatttca	cttttaatgt	aggcggcttg	tttgacaaca	3120
aaggcaattc	aaatatttcc	attgccaaag	gaggggctcg	ctttaagac	attgataatt	3180
ccaagaattt	aagcatcacc	accaactcca	gtccacttta	ccgcactatt	ataagcggca	3240
atataaccaa	taaaaacggg	gattttaaata	ttacgaacga	aggtagtgat	actgaaatgc	3300
aaattggcgg	cgatgtctcg	caaaaagaag	gtaatctcac	gatttcttct	gacaaaatca	3360
atattaccaa	acagataaca	atcaaggcag	gtgttgatgg	ggagaattcc	gattcagacg	3420
cgacaaacaa	tgccaatcta	accattaaaa	ccaagaat	gaaattaacg	caagacctaa	3480
atatttcagg	tttcaataaaa	gcagagatta	cagctaaaga	tggtagtgat	ttactatttg	3540
gtaacaccaa	tagtgctgat	ggtactaatg	ccaaaaaagt	aacctttaac	cagggttaaag	3600
attcaaaaat	ctctgctgac	ggtcaccaagg	tgacactaca	cagcaaagt	gaaacatccg	3660
gtagtaataa	caacactgaa	gatagcagtg	acaataatgc	cggcttaact	atcgatgcaa	3720
aaaatgtaac	agtaaacaac	aatattactt	ctcacaaagc	agtgcagcatc	tctgcgacaa	3780
gtggagaaat	taccactaaa	acaggtacaa	ccattaacgc	aaccactggg	aacgtggaga	3840
taaccgctca	aacaggtagt	atcctaggtg	gaattgagtc	cagctctggc	tctgtaacac	3900
ttactgcaac	cgagggcgct	cttgctgtaa	gcaatatttc	gggcaacacc	gttactgtta	3960
ctgcaaatag	cgggtgcatta	accacttttg	caggctctac	aattaaagga	accgagagt	4020
taaccacttc	aagtcaatca	ggcgatatcg	gcggtacgat	ttctgggtggc	acagtagagg	4080
ttaaagcaac	cgaaagttta	accactcaat	ccaattcaaa	aattaaagca	acaacaggcg	4140
aggctaacgt	aacaagtgca	acaggtacaa	ttggtggtac	gatttccggg	aatacggtaa	4200
atgttacggc	aaacgctggc	gatttaacag	ttgggaatgg	cgagaaaatt	aatgcgacag	4260
aaggagctgc	aaccttaact	acatcatcgg	gcaaattaac	taccgaagct	agttcacaca	4320
ttacttcagc	caagggtcag	gtaaatcttt	cagctcagga	tggtagcggt	gcaggaagta	4380
ttaatgccgc	caatgtgaca	ctaaatacta	caggcacttt	aactaccgtg	aagggttcaa	4440
acattaatgc	aaccagcggg	accttggtta	ttaacgcaaa	agacgctgag	ctaaatggcg	4500
cagcattggg	taaccacaca	gtggtaaatt	caaccaacgc	aaatggctcc	ggcagcgtaa	4560
tcgcgacaac	ctcaagcaga	gtgaacatca	ctggggattt	aatcacaata	aatggattaa	4620
atatcatttc	aaaaaacggg	ataaacaccg	tactgttaaa	aggcggttaa	attgatgtga	4680
aataacttca	accgggtata	gcaagcgtag	atgaagtaat	tgaagcgaaa	cgcatccttg	4740
agaaggtaaa	agatttatct	gatgaagaaa	gagtaagcgt	agctaaactt	ggagtaagt	4800
ctgtacgttt	tattgagcca	aataatacaa	ttacagtcga	tacacaaaaat	gaatttgcaa	4860
ccagaccatt	aagtcgaata	gtgatttctg	aaggcagggc	gtgtttctca	aacagtgatg	4920
gcgcgacggg	gtgcgttaat	atcgctgata	acgggcggta	gcggtcagta	attgacaagg	4980
tagatttcat	cctgcaatga	agtcatttta	ttttcgtatt	atttactgtg	tgggttaaag	5040
ttcagtagcg	gctttaccca	tcttgtaaaa	aattacggag	aatacaataa	agtattttta	5100
acaggttatt	atttatg					5116

<210> 2  
 <211> 1536  
 <212> PRT  
 <213> Haemophilus influenzae

<400> 2

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu  
 1 5 10 15

Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys  
 20 25 30

Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys  
 35 40 45

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln  
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr  
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val  
 85 90 95

Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
 100 105 110

Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val  
 115 120 125

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys  
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
 245 250 255

Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala  
 275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
 290 295 300  
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
 305 310 315 320  
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
 325 330 335  
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
 340 345 350  
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
 355 360 365  
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
 370 375 380  
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
 385 390 395 400  
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly  
 405 410 415  
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile  
 420 425 430  
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn  
 435 440 445  
 Ala Glu Thr Ala Gly Arg Ser Asn Thr Ser Glu Asp Asp Glu Tyr Thr  
 450 455 460  
 Gly Ser Gly Asn Ser Ala Ser Thr Pro Lys Arg Asn Lys Glu Lys Thr  
 465 470 475 480  
 Thr Leu Thr Asn Thr Thr Leu Glu Ser Ile Leu Lys Lys Gly Thr Phe  
 485 490 495  
 Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn  
 500 505 510  
 Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly  
 515 520 525  
 Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly  
 530 535 540  
 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn  
 545 550 555 560  
 Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp  
 565 570 575  
 Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr  
 580 585 590

Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu  
 595 600 605  
 Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys  
 610 615 620  
 Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys  
 625 630 635 640  
 Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys  
 645 650 655  
 Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu  
 660 665 670  
 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala  
 675 680 685  
 Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys  
 690 695 700  
 Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile  
 705 710 715 720  
 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser  
 725 730 735  
 Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Gly Ser Val Asp Phe Thr  
 740 745 750  
 Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn  
 755 760 765  
 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr  
 770 775 780  
 Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu  
 785 790 795 800  
 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp  
 805 810 815  
 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu  
 820 825 830  
 Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu  
 835 840 845  
 Gly Asn Val Thr Ile Asn Asn Asn Ala Asn Val Thr Leu Ile Gly Ser  
 850 855 860  
 Asp Phe Asp Asn His Gln Lys Pro Leu Thr Ile Lys Lys Asp Val Ile  
 865 870 875 880  
 Ile Asn Ser Gly Asn Leu Thr Ala Gly Gly Asn Ile Val Asn Ile Ala  
 885 890 895  
 Gly Asn Leu Thr Val Glu Ser Asn Ala Asn Phe Lys Ala Ile Thr Asn  
 900 905 910

Phe Thr Phe Asn Val Gly Gly Leu Phe Asp Asn Lys Gly Asn Ser Asn  
 915 920 925  
 Ile Ser Ile Ala Lys Gly Gly Ala Arg Phe Lys Asp Ile Asp Asn Ser  
 930 935 940  
 Lys Asn Leu Ser Ile Thr Thr Asn Ser Ser Ser Thr Tyr Arg Thr Ile  
 945 950 955 960  
 Ile Ser Gly Asn Ile Thr Asn Lys Asn Gly Asp Leu Asn Ile Thr Asn  
 965 970 975  
 Glu Gly Ser Asp Thr Glu Met Gln Ile Gly Gly Asp Val Ser Gln Lys  
 980 985 990  
 Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr Lys Gln  
 995 1000 1005  
 Ile Thr Ile Lys Ala Gly Val Asp Gly Glu Asn Ser Asp Ser Asp Ala  
 1010 1015 1020  
 Thr Asn Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys Leu Thr  
 1025 1030 1035 1040  
 Gln Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala Lys  
 1045 1050 1055  
 Asp Gly Ser Asp Leu Thr Ile Gly Asn Thr Asn Ser Ala Asp Gly Thr  
 1060 1065 1070  
 Asn Ala Lys Lys Val Thr Phe Asn Gln Val Lys Asp Ser Lys Ile Ser  
 1075 1080 1085  
 Ala Asp Gly His Lys Val Thr Leu His Ser Lys Val Glu Thr Ser Gly  
 1090 1095 1100  
 Ser Asn Asn Asn Thr Glu Asp Ser Ser Asp Asn Asn Ala Gly Leu Thr  
 1105 1110 1115 1120  
 Ile Asp Ala Lys Asn Val Thr Val Asn Asn Asn Ile Thr Ser His Lys  
 1125 1130 1135  
 Ala Val Ser Ile Ser Ala Thr Ser Gly Glu Ile Thr Thr Lys Thr Gly  
 1140 1145 1150  
 Thr Thr Ile Asn Ala Thr Thr Gly Asn Val Glu Ile Thr Ala Gln Thr  
 1155 1160 1165  
 Gly Ser Ile Leu Gly Gly Ile Glu Ser Ser Ser Gly Ser Val Thr Leu  
 1170 1175 1180  
 Thr Ala Thr Glu Gly Ala Leu Ala Val Ser Asn Ile Ser Gly Asn Thr  
 1185 1190 1195 1200  
 Val Thr Val Thr Ala Asn Ser Gly Ala Leu Thr Thr Leu Ala Gly Ser  
 1205 1210 1215

Thr Ile Lys Gly Thr Glu Ser Val Thr Thr Ser Ser Gln Ser Gly Asp  
 1220 1225 1230  
 Ile Gly Gly Thr Ile Ser Gly Gly Thr Val Glu Val Lys Ala Thr Glu  
 1235 1240 1245  
 Ser Leu Thr Thr Gln Ser Asn Ser Lys Ile Lys Ala Thr Thr Gly Glu  
 1250 1255 1260  
 Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly Thr Ile Ser Gly  
 1265 1270 1275 1280  
 Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu Thr Val Gly Asn  
 1285 1290 1295  
 Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr Leu Thr Thr Ser  
 1300 1305 1310  
 Ser Gly Lys Leu Thr Thr Glu Ala Ser Ser His Ile Thr Ser Ala Lys  
 1315 1320 1325  
 Gly Gln Val Asn Leu Ser Ala Gln Asp Gly Ser Val Ala Gly Ser Ile  
 1330 1335 1340  
 Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Val  
 1345 1350 1355 1360  
 Lys Gly Ser Asn Ile Asn Ala Thr Ser Gly Thr Leu Val Ile Asn Ala  
 1365 1370 1375  
 Lys Asp Ala Glu Leu Asn Gly Ala Ala Leu Gly Asn His Thr Val Val  
 1380 1385 1390  
 Asn Ala Thr Asn Ala Asn Gly Ser Gly Ser Val Ile Ala Thr Thr Ser  
 1395 1400 1405  
 Ser Arg Val Asn Ile Thr Gly Asp Leu Ile Thr Ile Asn Gly Leu Asn  
 1410 1415 1420  
 Ile Ile Ser Lys Asn Gly Ile Asn Thr Val Leu Leu Lys Gly Val Lys  
 1425 1430 1435 1440  
 Ile Asp Val Lys Tyr Ile Gln Pro Gly Ile Ala Ser Val Asp Glu Val  
 1445 1450 1455  
 Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp Leu Ser Asp Glu  
 1460 1465 1470  
 Glu Arg Glu Ala Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Ile  
 1475 1480 1485  
 Glu Pro Asn Asn Thr Ile Thr Val Asp Thr Gln Asn Glu Phe Ala Thr  
 1490 1495 1500  
 Arg Pro Leu Ser Arg Ile Val Ile Ser Glu Gly Arg Ala Cys Phe Ser  
 1505 1510 1515 1520  
 Asn Ser Asp Gly Ala Thr Val Cys Val Asn Ile Ala Asp Asn Gly Arg  
 1525 1530 1535

<210> 3  
 <211> 4937  
 <212> DNA  
 <213> Haemophilus influenzae

<400> 3  
 taaatataca agataataaaa aataaatcaa gattttttgtg atgacaaaaca acaattacaa 60  
 cacctttttt gcagtctata tgcaaatatt ttaaaaaaat agtataaatc cgccatataa 120  
 aatgggtataa tctttcatct ttcattcttta atctttcatc tttcatcttt catctttcat 180  
 ctttcatctt tcatctttca tctttcatct ttcattcttc atctttcatc tttcatcttt 240  
 cacatgaaat gatgaaccga gggaagggag ggaggggcaa gaatgaagag ggagctgaac 300  
 gaacgcaaata gataaagtaa ttttaattgtt caactaacct taggagaaaa tatgaacaag 360  
 atatatcgtc tcaaattcag caaacgcctg aatgcttttg ttgctgtgtc tgaattggca 420  
 cgggggtgtg accattccac agaaaaaggc ttccgctatg ttactatctt taggtgtaac 480  
 cacttagcgt taaagccact ttccgctatg ttactatctt taggtgtaac atctattcca 540  
 caatctgttt tagcaagcgg cttacaagga atggatgtag tacacggcac agccactatg 600  
 caagtagatg gtaataaaaac cattatccgc aacagtgttg acgctatcat taattggaaa 660  
 caatttaaca tcgacaaaaa tgaaatgggtg cagtttttac aagaaaacaa caactccgcc 720  
 gtattcaacc gtgttacatc taaccaaatac tcccaattaa aagggatttt agattctaac 780  
 ggacaagtct ttttaatacaa cccaaatggg atcacaatag gtaaagacgc aattattaac 840  
 actaatggct ttacggcttc tacgctagac atttctaacg aaaacatcaa ggcgcgtaac 900  
 ttcaccttcg agcaaaccaa agataaagcg ctgcgtgaaa ttgtgaatca cgggttaatt 960  
 actgtcggta aagacggcag tgtaaatctt attggtggca aagtgaaaaa cgaggggtgtg 1020  
 attagcgtaa atggtggcag catttcttta ctgcagggc aaaaaatcac catcagcgat 1080  
 ataataaacc caaccattac ttacagcatt gccgcgcctg aaaaatgaagc ggtcaatctg 1140  
 ggcgatattt ttgccaaagg cggtaacatt aatgtccgtg ctgccactat tcgaaaccaa 1200  
 ggtaaacctt ctgctgattc tgtaagcaaa gataaaagcg gcaatattgt tctttccgcc 1260  
 aaagaggggtg aagcggaaat tggcgggtgta atttccgctc aaaaatcagca agctaaaggc 1320  
 ggcaagctga tgattacagg cgataaagtc acattaaaaa caggtgcagt tatcgacctt 1380  
 tcaggtaaaag aaggggggaga aacttacctt ggcggtgacg agcgcggcga aggtaaaaac 1440  
 ggcatccaat tagcaaaagaa aacctcttta gaaaaaggct caaccatcaa tgtatcaggc 1500  
 aaagaaaaag gcgagcgcgc tattgtgtgg ggcgatattg cgtaattga cggcaatatt 1560  
 aacgctcaag gtagtggtga tatcgctaaa accggtggtt ttgtggagac atcggggcat 1620  
 tatttatcca ttgacagcaa tgcaattggt aaaaacaaaag agtggttgct agaccctgat 1680  
 gatgtaacaa ttgaagccga agaccccctt cgaataata ccggtataaa tgatgaattc 1740  
 ccaacaggca ccggtgaagc aagcgaccct aaaaaaata gcgaactcaa aacaacgcta 1800  
 accaatacaa ctatttcaaa ttatctgaaa aacgcctgga caatgaatat aacggcatca 1860  
 agaaaactta ccgttaatag ctcaatcaac atcggaagca actcccactt aattctccat 1920  
 agtaaaggct agcgtggcgg aggcgttcag attgatggag atattacttc taaaggcgga 1980  
 aatttaacca tttattctgg cggtatgggtt gatgttcata aaaatattac gcttgatcag 2040  
 ggttttttaa atattaccgc cgcttccgta gcttttgaag gtggaataa caaagcacgc 2100  
 gacgcggcaa atgctaaaat tgtcgcccag ggcaactgaa ccattacagg agaggggaaa 2160  
 gatttcaggg ctaacaacgt atcttttaac ggaacgggtg aaggtctgaa tatcatttca 2220  
 tcagtgaata atttaaccca caatcttagt ggcacaatta acatatctgg gaataaaca 2280  
 attaaccaaa ctacgagaaa gaacacctcg tattggcaaa ccagccatga ttgcactgg 2340  
 aacgtcagtg ctcttaatct agagacaggc gcaaatttta cctttattaa atacatttca 2400  
 agcaatagca aaggcttaac aacacagtat agaagctctg caggggtgaa ttttaacggc 2460  
 gtaaattggca acatgtcatt caatctcaaa gaaggagcga aagttaattt caaattaaaa 2520  
 ccaaacgaga acatgaacac aagcaaacct ttaccaattc ggtttttagc caatatcaca 2580  
 gccactgggt ggggctctgt tttttttgat atatatgcca accattctgg cagaggggct 2640  
 gagttaaaaa tgagtgaat taatatctct aacggcgcta atttacctt aaattcccat 2700  
 gttcgcggcg atgacgcttt taaaatcaac aaagacttaa ccataaatgc aaccaattca 2760  
 aatttcagcc tcagacagac gaaagatgat ttttatgacg ggtacgcacg caatgccatc 2820  
 aattcaacct acaacatatc cattctgggc ggtaatgtca cccttgggtg acaaaactca 2880



```

agcagcagca ttacggggaa tattactatc gagaaagcag caaatgttac gctagaagcc 2940
aataacgccc ctaatcagca aaacataagg gatagagtta taaaacttgg cagcttgctc 3000
gttaattggga gtttaagttt aactggcgaa aatgcagata ttaaaggcaa tctcactatt 3060
tcagaaagcg ccacttttaa aggaaagact agagataccc taaatatcac cggcaatttt 3120
accaataatg gcactgccga aattaatata acacaaggag tggtaaaact tggcaatgtt 3180
accaatgatg gtgatttaaa cattaccact cagcgtaaac gcaaccaaag aagcatcatc 3240
ggcggagata taatcaacaa aaaaggaagc ttaaattatta cagacagtaa taatgatgct 3300
gaaatccaaa ttggcggcaa tatctcgcaa aaagaaggca acctcacgat ttcttcgat 3360
aaaattaata tcaccaaaaca gataacaatc aaaaagggtta ttgatggaga ggactctagt 3420
tcagatgcga caagtaatgc caacctaaact attaaaacca aagaattgaa attgacagaa 3480
gacctaagta tttcagggtt caataaagca gagattacag ccaaagatgg tagagattta 3540
actattggca acagtaatga cggtaacagc ggtgccgaag ccaaaacagt aacttttaac 3600
aatgttaag attcaaaaat ctctgctgac ggtcacaatg tgacactaaa tagcaaagt 3660
aaaacatcta gcagcaatgg cggacgtgaa agcaatagcg acaacgatac cggcttaact 3720
attactgcaa aaaatgtaga agtaaacaaa gatattactt ctctcaaaac agtaaataatc 3780
accgcgtcgg aaaagggttac caccacagca ggctcgacca ttaacgcaac aaatggcaaa 3840
gcaagtatta caaccaaacc aggtgatatc agcggtagca tttccggtaa cagggttaagt 3900
gttagcgcga ctgggtgattt aaccactaaa tccggctcaa aaattgaagc gaaatcgggt 3960
gaggctaattg taacaagtgc aacagggtaca attggcggtta caatttcgg taatacggta 4020
aatgttacgg caaacgctgg cgatttaaca gttgggaatg gcgcagaaat taatgcgaca 4080
gaaggagctg caaccttaac cgcaacaggg aataccttga ctactgaagc cgttcttagc 4140
atcacttcaa ctaagggtca ggtagacctc ttggctcaga atggtagcat cgcaggaagc 4200
attaatgctg ctaatgtgac attaaatact acaggcacct taaccaccgt gccaggctcg 4260
gatattaaag caaccagcgg caccttggtt attaacgcaa aagatgctaa gctaaatggt 4320
gatgcaagg gtgatagtac agaagtgaat gcagtcaacg caagcggctc tggtagtgtg 4380
actgcggcga cctcaagcag tgtgaatatc actggggatt taaacacagt aaatgggtta 4440
aatatcattt cgaaagatgg tagaaacact gtgcgcttaa gaggcaagga aattgagggtg 4500
aaatatatcc agccagggtg agcaagtgtg gaagaagtaa ttgaagcgaa acgcgtcctt 4560
gaaaaagtaa aagattttatc tgatgaagaa agagaaacat tagctaaact tgggtgaagt 4620
gctgtacgtt ttgttgagcc aaataatata attacagtca atacacaaaa tgaatttaca 4680
accagaccgt caagtcaggt gataatttct gaaggtaagg cgtgtttctc aagtggtaat 4740
ggcgacagag tatgtaccaa tgttgctgac gatggacagc cgtagtcagt aattgacaag 4800
gtagatttca tcctgcaatg aagtcatttt attttcgtat tatttactgt gtgggttaaa 4860
gttcagtacg ggctttaccc atcttgtaaa aaattacgga gaatacaata aagtattttt 4920
aacaggttat tattatg 4937

```

&lt;210&gt; 4

&lt;211&gt; 1477

&lt;212&gt; PRT

&lt;213&gt; Haemophilus influenzae

&lt;400&gt; 4

```

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu
  1             5             10             15

```

```

Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys
          20             25             30

```

```

Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys
          35             40             45

```

```

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
          50             55             60

```

```

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
          65             70             75             80

```

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val  
 85 90 95  
 Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
 100 105 110  
 Val Gln Phe Leu Gln Glu Asn Asn Asn Ser Ala Val Phe Asn Arg Val  
 115 120 125  
 Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
 130 135 140  
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
 145 150 155 160  
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
 165 170 175  
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys  
 180 185 190  
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
 195 200 205  
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
 210 215 220  
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
 225 230 235 240  
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
 245 250 255  
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
 260 265 270  
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala  
 275 280 285  
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
 290 295 300  
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
 305 310 315 320  
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
 325 330 335  
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
 340 345 350  
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
 355 360 365  
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
 370 375 380  
 Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
 385 390 395 400

Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly  
 405 410 415  
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile  
 420 425 430  
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn  
 435 440 445  
 Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro  
 450 455 460  
 Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys  
 465 470 475 480  
 Thr Thr Leu Thr Asn Thr Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp  
 485 490 495  
 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile  
 500 505 510  
 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg  
 515 520 525  
 Gly Gly Gly Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Gly Asn  
 530 535 540  
 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr  
 545 550 555 560  
 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu  
 565 570 575  
 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala  
 580 585 590  
 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn  
 595 600 605  
 Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser  
 610 615 620  
 Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly  
 625 630 635 640  
 Asn Ile Thr Ile Asn Gln Thr Thr Arg Lys Asn Thr Ser Tyr Trp Gln  
 645 650 655  
 Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr  
 660 665 670  
 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly  
 675 680 685  
 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val  
 690 695 700

Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Gly Ala Lys Val Asn Phe  
 705 710 715 720  
 Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile  
 725 730 735  
 Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Gly Ser Val Phe Phe  
 740 745 750  
 Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser  
 755 760 765  
 Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val  
 770 775 780  
 Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala  
 785 790 795 800  
 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp  
 805 810 815  
 Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu  
 820 825 830  
 Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ser Ile Thr  
 835 840 845  
 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn  
 850 855 860  
 Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly  
 865 870 875 880  
 Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp  
 885 890 895  
 Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys  
 900 905 910  
 Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr  
 915 920 925  
 Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr  
 930 935 940  
 Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg  
 945 950 955 960  
 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile  
 965 970 975  
 Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser  
 980 985 990  
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr  
 995 1000 1005  
 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser  
 1010 1015 1020

Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys  
 1025 1030 1035 1040  
 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr  
 1045 1050 1055  
 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn  
 1060 1065 1070  
 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser  
 1075 1080 1085  
 Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys  
 1090 1095 1100  
 Thr Ser Ser Ser Asn Gly Gly Arg Glu Ser Asn Ser Asp Asn Asp Thr  
 1105 1110 1115 1120  
 Gly Leu Thr Ile Thr Ala Lys Asn Val Glu Val Asn Lys Asp Ile Thr  
 1125 1130 1135  
 Ser Leu Lys Thr Val Asn Ile Thr Ala Ser Glu Lys Val Thr Thr Thr  
 1140 1145 1150  
 Ala Gly Ser Thr Ile Asn Ala Thr Asn Gly Lys Ala Ser Ile Thr Thr  
 1155 1160 1165  
 Lys Thr Gly Asp Ile Ser Gly Thr Ile Ser Gly Asn Thr Val Ser Val  
 1170 1175 1180  
 Ser Ala Thr Val Asp Leu Thr Thr Lys Ser Gly Ser Lys Ile Glu Ala  
 1185 1190 1195 1200  
 Lys Ser Gly Glu Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly  
 1205 1210 1215  
 Thr Ile Ser Gly Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu  
 1220 1225 1230  
 Thr Val Gly Asn Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr  
 1235 1240 1245  
 Leu Thr Ala Thr Gly Asn Thr Leu Thr Thr Glu Ala Gly Ser Ser Ile  
 1250 1255 1260  
 Thr Ser Thr Lys Gly Gln Val Asp Leu Leu Ala Gln Asn Gly Ser Ile  
 1265 1270 1275 1280  
 Ala Gly Ser Ile Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr  
 1285 1290 1295  
 Leu Thr Thr Val Ala Gly Ser Asp Ile Lys Ala Thr Ser Gly Thr Leu  
 1300 1305 1310  
 Val Ile Asn Ala Lys Asp Ala Lys Leu Asn Gly Asp Ala Ser Gly Asp  
 1315 1320 1325

Ser Thr Glu Val Asn Ala Val Asn Ala Ser Gly Ser Gly Ser Val Thr  
 1330 1335 1340

Ala Ala Thr Ser Ser Ser Val Asn Ile Thr Gly Asp Leu Asn Thr Val  
 1345 1350 1355 1360

Asn Gly Leu Asn Ile Ile Ser Lys Asp Gly Arg Asn Thr Val Arg Leu  
 1365 1370 1375

Arg Gly Lys Glu Ile Glu Val Lys Tyr Ile Gln Pro Gly Val Ala Ser  
 1380 1385 1390

Val Glu Glu Val Ile Glu Ala Lys Arg Val Leu Glu Lys Val Lys Asp  
 1395 1400 1405

Leu Ser Asp Glu Glu Arg Glu Thr Leu Ala Lys Leu Gly Val Ser Ala  
 1410 1415 1420

Val Arg Phe Val Glu Pro Asn Asn Thr Ile Thr Val Asn Thr Gln Asn  
 1425 1430 1435 1440

Glu Phe Thr Thr Arg Pro Ser Ser Gln Val Ile Ile Ser Glu Gly Lys  
 1445 1450 1455

Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala  
 1460 1465 1470

Asp Asp Gly Gln Pro  
 1475

<210> 5

<211> 9171

<212> DNA

<213> Haemophilus influenzae

<400> 5

acagcggttct cttaataacta gtacaaaccc acaataaaaat atgacaaaaca acaattacaa 60  
 caccttttttt gcagtctata tgcaaatatt ttaaaaaata gtataaatcc gccatataaa 120  
 atggtataat ctttcatctt tcatctttca tctttcatct ttcattcttc atctttcatc 180  
 tttcatcttt catctttcat ctttcatctt tcatctttca tctttcatct ttcattcttc 240  
 acatgaaatg atgaaccgag ggaagggagg gaggggcaag aatgaagagg gagctgaacg 300  
 aacgcaaatg ataaagtaat ttaattgttc aactaacctt aggagaaaat atgaacaaga 360  
 tatatcgctc caaattcagc aaacgcctga atgctttggg tgctgtgtct gaattggcac 420  
 ggggttggtga ccattccaca gaaaaaggca gcgaaaaacc tgctcgcatg aaagtgcgtc 480  
 acttagcggt aaagccactt tccgctatgt tactatcttt aggtgtaaca tctattccac 540  
 aatctgtttt agcaagcggc ttacaaggaa tggatgtagt acacggcaca gccactatgc 600  
 aagtagatgg taataaaacc attatccgca acagtgttga cgctatcatt aattggaaac 660  
 aatttaacat cgacaaaaat gaaatggtgc agttttttaca agaaaacaac aactccgccg 720  
 tattcaaccg tgttacatct aaccaaactc cccaattaaa agggatttta gattctaacg 780  
 gacaagtctt tttaatcaac ccaaaggta tcacaatagg taaagacgca attattaaca 840  
 ctaatggctt tacggcttct acgctagaca tttctaacga aaacatcaag gcgcgtaatt 900  
 tcaccttcga gcaaaccaaa gataaagcgc tcgctgaaat tgtgaatcac ggtttaatta 960  
 ctgtcggtaa agacggcagt gtaaatctta ttggtggcaa agtgaaaaac gaggggtgtga 1020  
 ttagcgtaaa tgggtggcagc atttctttac tcgcagggca aaaaatcacc atcagcgata 1080  
 taataaaccc aaccattact tacagcattg ccgcgctga aaatgaagcg gtcaatctgg 1140  
 gcgatatctt tgccaaaggc ggtaacatta atgtccgtgc tgccactatt cgaaaccaag 1200  
 ctttccgccca aagagggtga agcggaaatt ggcggtgtaa tttccgctca aaatcagcaa 1260  
 gctaaaggcg gcaagctgat gattacagcg gataaagtca cattaaaaac aggtgcagtt 1320

atcgaccttt	caggtaaaga	agggggagaa	acttaccttg	gcggtgacga	gcgcggcgaa	1380
ggtaaaaacg	gcattcaatt	agcaaagaaa	acctcttttag	aaaaaggctc	aaccatcaat	1440
gtatcaggca	aagaaaaagg	cggacgcgct	attgtgtggg	gcgatattgc	gttaattgac	1500
ggcaatatta	acgctcaagg	tagtggtgat	atcgctaaaa	cgggtggttt	tgtggagacg	1560
tcggggcatg	atttattcat	caaagacaat	gcaattgttg	acgccaaaga	gtggttggtta	1620
gacccggata	atgtatctat	taatgcagaa	acagcaggac	gcagcaatac	ttcagaagac	1680
gatgaatata	cgggatccgg	gaatagtgcc	agcaccctaa	aacgaaacaa	agaaaagaca	1740
acattaacaa	acacaactct	tgagagtata	ctaaaaaaaag	gtacctttgt	taacatcact	1800
gctaatacaac	gcatctatgt	caatagctcc	attaattttat	ccaatggcag	cttaactctt	1860
tggagtggagg	gtcggagcgg	tggcggcggt	gagattaaca	acgatattac	caccggtgat	1920
gataccagag	gtgcaaactt	aacaattttac	tcaggcggct	gggttgatgt	tcataaaaaat	1980
atctcactcg	gggcgcaagg	taacataaac	attacagcta	aacaagatat	cgcctttgag	2040
aaagggaagca	accaagtcat	tacaggtcaa	gggactatta	cctcaggcaa	tcaaaaagggt	2100
tttagatttta	ataatgtctc	tctaaacggc	actggcagcg	gactgcaatt	caccactaaa	2160
agaaccaata	aatacgctat	cacaaataaaa	tttgaaggga	ctttaaatat	ttcagggaaa	2220
gtgaacatct	caatggtttt	acctaaaaat	gaaagtggat	atgataaatt	caaaggacgc	2280
acttactgga	atttaacctc	gaaagtggat	atgataaatt	caaaggacgc	cctcactatt	2340
gactccagag	gaagcgatag	tgacggcaca	cttaccagc	cttataattt	aaacggtata	2400
tcattcaaca	aagacactac	ctttaatggt	gaacgaaatg	caagagtcaa	ctttgacatc	2460
aaggcaccaa	tagggataaaa	taagtattct	agtttgaatt	acgcattcatt	taatggaaac	2520
atttcagttt	cgggaggggg	gagtgttgat	ttcacacttc	tcgcctcatc	ctctaacgtc	2580
caaaccctcg	gtgtagttat	aaattctaaa	tacttttaatg	tttcaacagg	gtcaagttta	2640
agatttaaaa	cttcaggctc	aacaaaaact	ggcttctcaa	tagagaaaga	tttaacttta	2700
aatgccacog	gaggcaacat	aacacttttg	caagttgaag	gcaccgatgg	aatgattggg	2760
aaaggcattg	tagccaaaaa	aaacataacc	tttgaaggag	gtaagatgag	gtttggctcc	2820
aggaaagccg	taacagaaat	cgaaggcaat	gttactatca	ataacaacgc	taacgtcact	2880
cttatcggtt	cggattttga	caaccatcaa	aaacctttta	ctattaaaaa	agatgtcatc	2940
attaatagcg	gcaaccttac	cgtcggaggc	aatatgttca	atatagcggg	aaatcttacc	3000
gttgaaagta	acgctaattt	caaagctatc	acaaatttca	cttttaatgt	aggcggcttg	3060
tttgacaaca	aaggcaattc	aaatatttcc	attgccaaag	gaggggctcg	ctttaaagac	3120
attgataatt	ccaagaattt	aagcatcacc	accaactcca	gctccactta	ccgcactatt	3180
ataagcggca	atataaccaa	taaaaacggg	gatttaataa	ttacgaacga	aggtagtgat	3240
actgaaatgc	aaattggcgg	cgatgtctcg	caaaaagaag	gtaatctcac	gatttcttct	3300
gacaaaatca	atattaccaa	acagataaca	atcaaggcag	gtgttgatgg	ggagaattcc	3360
gattcagacg	cgacaaaaca	tgccaatcta	accattaaaa	ccaaagaatt	gaaattaacg	3420
caagacctaa	atatttcagg	tttcaataaaa	gcagagatta	cagctaaaga	tggtagtgat	3480
ttaactattg	gtaacaccaa	tagtgctgat	ggtactaatg	ccaaaaaagt	aacctttaac	3540
cagggttaaag	attcaaaaat	ctctgctgac	ggtcacagg	tgacactaca	cagcaaagtg	3600
gaaacatcog	gtagtaataa	caacactgaa	gatagcagtg	acaataatgc	cggcttaact	3660
atcgatgcaa	aaaatgtaac	agtaaacaac	aatattactt	ctcacaaagc	agtgagcatc	3720
tctgcgacaa	gtggagaaat	taccactaaa	acaggtacaa	ccattaacgc	aaccactggg	3780
aacgtggaga	taaccgctca	aacaggtagt	atcctagggtg	gaattgagtc	cagctctggc	3840
tctgtaacac	ttactgcaac	cgagggcgct	cttgctgtaa	gcaatatttc	gggcaacacc	3900
gttactgtta	ctgcaaatag	cgggtgcatta	accactttgg	caggctctac	aattaaagga	3960
accgagagtg	taaccacttc	aagtcaatca	ggcgatatcg	gcggtacgat	ttctgggtggc	4020
acagtagagg	ttaaagcaac	cgaaggttta	accactcaat	ccaattcaaa	aattaaagca	4080
acaacaggcg	aggctaacgt	aacaagtgc	acaggtacaa	ttggtggtac	gatttccggg	4140
aatacggtaa	atgttacggc	aaacgctggc	gatttaacag	ttgggaatgg	cgcagaaatt	4200
aatgcgacag	aaggagctgc	aaccttaact	acatcatcgg	gcaaattaac	taccgaagct	4260
agttcacaca	ttacttcagc	caagggtcag	gtaaactctt	cagctcagga	tggtagcgtt	4320
gcaggaagta	ttaatgccgc	caatgtgaca	ctaaatacta	caggcacttt	aactaccgtg	4380
aagggttcaa	acattaatgc	aaccagcggg	accttggtta	ttaacgcaaa	agacgctgag	4440
ctaaatggcg	cagcattggg	taaccacaca	gtggtaaaatg	caaccaacgc	aaatggctcc	4500
ggcagcgtaa	tcgcgcacaac	ctcaagcaga	gtgaacatca	ctggggattt	aatcacaata	4560
aatggattaa	atatcatttc	aaaaaacggg	ataaacaccg	tactgttaaa	aggcggttaa	4620
attgatgtga	aatacattca	accgggtata	gcaagcgtag	atgaagtaat	tgaagcgaaa	4680
cgcactcttg	agaaggtaaa	agatttatct	gtagaagaaa	gagaagcgtt	agctaaactt	4740
ggcgtaagtg	ctgtacggtt	tattgagcca	aataatacaa	ttacagtcga	tacacaaaat	4800
gaatttgcaa	ccagaccatt	aagtccaata	gtgatttctg	aaggcagggc	gtgtttctca	4860

aacagtgatg	gcgcgacggg	gtgctgtaat	atcgctgata	acggggcggt	gcggtcagta	4920
attgacaagg	tagatttcat	cctgcaatga	agtcatttta	ttttcgtatt	atttactgtg	4980
tgggttaaag	ttcagtagcg	gctttaccca	tcttgtaaaa	aattacggag	aatacaataa	5040
agtattttta	acaggttatt	attatgaaaa	atataaaaaag	cagattaaaa	ctcagtgcaa	5100
tatcagttat	gcttggcctg	gcttcttcat	cattgtatgc	agaagaagcg	tttttagtaa	5160
aaggctttca	gttatctggg	gcacttgaaa	ctttaagtga	agacgccccaa	ctgtctgtag	5220
caaaatcttt	atctaaatac	caaggctcgc	aaactttaac	aaacctaaaa	acagcacagc	5280
ttgaattaca	ggctgtgcta	gataagattg	agccaaataa	gtttgatgtg	atattgccac	5340
aacaaaccaa	tacggatggc	aatattatgt	ttgagctagt	ctcgaaatca	gccgcagaaa	5400
gccaaagttaa	ttataaggcg	agccagggtt	atagtgaaga	aaatatcgct	cgtagcctgc	5460
catctttgaa	acaaggaaaa	gtgtatgaag	atggctcgta	gtggttcgat	ttgcgtgaat	5520
tcaatatggc	aaaagaaaaa	ccacttaaag	tcactcgcgt	gcattacgag	ttaaacctta	5580
aaaacaaaac	ctctgatttg	gtagttagcag	gtttttcgcc	ttttggcaaa	acgcgtagct	5640
ttgtttccta	tgataatttc	ggcgcaaggg	agtttaacta	tcaacgtgta	agtcagggtt	5700
ttgtaaatgc	caatttgacc	ggacatgatg	atgtattaaa	tctaaacgca	ttgaccaatg	5760
taaaagcacc	atcaaaatct	tatgcggtag	gcataaggata	tacttatccg	ttttatgata	5820
aacaccaatc	cttaagtctt	tataccagca	tgagttatgc	tgattctaata	gatatcgacg	5880
gcttaccagg	tgcgattaat	cgtaaattat	caaaagggtca	atctatctct	gcgaatctga	5940
aatggaggtta	ttatctcccg	acatttaacc	ttggaatgga	agaccagttt	aaaatttaatt	6000
taggtacaaa	ctaccgcat	attaatcaaa	catccgagtt	aaacaccctg	ggtgcaacga	6060
agaaaaaatt	tgtagtatca	ggcgtaagtg	caggcattga	tggaacatac	caattttacc	6120
ctaaaaaat	ctttaatatt	gatttaactc	atcattatta	cgcgagtaaa	ttaccaggct	6180
cttttggaat	ggagcgcat	ggcgaaacat	ttaatcgtag	ctatcacatt	agcacagcca	6240
gtttagggtt	gagtagaag	tttgctcaag	gttggtcatt	tagcagtagc	ttatcggtgc	6300
agtttactct	acaagatata	agtagcatag	atttattctc	tgtaacagggt	acttatggcg	6360
tcagaggctt	taaatacggc	ggtgcaagtg	gtgagcgcg	tcttgtagtg	cgtaatgaat	6420
taagtatgcc	aaaatacacc	cgttttcaaa	tcagccctta	tgcgttttat	gatgcagggtc	6480
agttccggtta	taatagcgaa	aatgctaaaa	cttacggcga	agatatgcac	acggtatcct	6540
ctgcgggttt	aggcattaaa	acctctccta	cacaaaactt	aagcttagat	gcttttggtg	6600
ctcgtcgctt	tgcaaatgcc	aatagtgaac	atttgaatgg	caacaaaaaa	cgcacaaagt	6660
cacctacaa	ctctggggg	agattaacat	tcagtttcta	acctgaaat	ttaatcaact	6720
ggtaagcggt	ccgcctacca	gtttataact	atatgcttta	cccgccaatt	tacagtctat	6780
acgcaaccct	gttttcatcc	ttatatatca	aacaaactaa	gcaaaccaag	caaaccaagc	6840
aaaccaagca	aaccaagcaa	accaagcaaa	ccaagcaaac	caagcaaac	aagcaaacca	6900
agcaaaccaa	gcaaaccaag	caaaccaagc	aaaccaagca	atgctaaaaa	acaatttata	6960
tgataaaacta	aaacatactc	cataccatgg	caatacaagg	gatttaataa	tatgacaaaa	7020
gaaaatttac	aaagtgttcc	acaaaatacg	accgcttcac	ttgtagaatc	aaacaacgac	7080
caaacttccc	tgcaaatact	taaaacaacca	cccaaaccga	acctattacg	cctggaacaa	7140
catgtcgcca	aaaaagatta	tgagcttgct	tgccgcgaat	taatggcgat	tttggaaaaa	7200
atggacgcta	attttgagg	cgttcacgat	attgaatttg	acgcacctgc	tcagctggca	7260
tatctaccog	aaaaactact	aattcatttt	gccactcgtc	tcgctaattgc	aattacaaca	7320
ctcttttccg	accccgatt	ggcaatttcc	gaagaagggg	cattaaagat	gattagcctg	7380
caacgctggt	tgacgctgat	ttttgcctct	tccccctacg	ttaacgcaga	ccatattctc	7440
aataaatata	atatcaaccc	agattccgaa	ggtggctttc	atttagcaac	agacaactct	7500
tctattgcta	aattctgtat	tttttactta	cccgaatcca	atgtcaatat	gagtttagat	7560
gcgttatggg	cagggaatca	acaactttgt	gcttcattgt	gttttgctgt	gcagtcctca	7620
cgttttattg	gtactgcatc	tgcgtttcat	aaaagagcgg	tggttttaca	gtggtttctc	7680
aaaaaactcg	ccgaaattgc	taatttagat	gaattgcctg	caaatatcct	tcagtagtga	7740
tatatgcact	gcagttatga	tttagcaaaa	aacaagcacg	atgtaagcg	tccattaaac	7800
gaacttgctc	gcaagcatat	cctcacgcaa	ggatggcaag	accgctacct	ttacacctta	7860
ggtaaaaagg	acggcaaac	tgtagtagat	gtactgcttg	aacattttta	ttcgggacat	7920
tcgatttatc	gcacgcattc	aacttcaatg	attgctgctc	gagaaaaatt	ctatttagtc	7980
ggcttaggcc	atgagggcgt	tgataacata	ggtcgagaag	tgtttgacga	gttctttgaa	8040
atcagtagca	ataatataat	ggagagactg	ttttttatcc	gtaaacagtg	cgaaactttc	8100
caaccgcgag	tgttctatat	gccaagcatt	ggcatggata	ttaccacgat	ttttgtgagc	8160
aacactcggc	ttgcccctat	tcaagctgta	gccttgggtc	atcctgccac	tacgcattct	8220
gaattttattg	attatgtcat	cgtagaagat	gattatgtgg	gcagtgaaaga	ttgttttagc	8280
gaaacccttt	tacgcttacc	caaagatgcc	ctaccttatg	taccatctgc	actcgcccca	8340
caaaaagtgg	attatgtact	cagggaatac	cctgaagtag	tcaatatcgg	tattgccgct	8400



```

accacaatga aattaaaccc tgaatttttg ctaacattgc aagaaatcag agataaagct 8460
aaagtcaaaa tacattttca tttcgactt ggacaatcaa caggcttgac acacccttat 8520
gtcaaatggt ttatcgaaag ctatttaggt gacgatgcca ctgcacatcc ccacgcacct 8580
tatcacgatt atctggcaat attgctgat tgcgatatgc tactaaatcc gtttcctttc 8640
ggtaatacta acggcataat tgatatggtt acattaggtt tagttggtgt atgcaaaacg 8700
ggggatgaag tacatgaaca tattgatgaa ggtctgttta aacgcttagg actaccagaa 8760
tggctgatag ccgacacacg agaaacatat attgaatgtg ctttgcgtct agcagaaaaac 8820
catcaagaac gccttgaact ccgtcgttac atcatagaaa acaacggctt acaaaagctt 8880
tttacaggcg accctcgtcc attgggcaaa atactgctta agaaaacaaa tgaatggaag 8940
cggaagcact tgagtaaaaa ataacggttt tttaaagtaa aagtgcggtt aattttcaaa 9000
gcgttttaaa aacctctcaa aaatcaaccg cacttttatc tttataacgc tcccgcgcgc 9060
tgacagttaa tctctttctt aaaataacca taaaattgtg gcaatagttg ggtaatcaaa 9120
ttcaattgtt gatacggcaa actaaagacg gcgcgttctt cggcagtcac c 9171

```

&lt;210&gt; 6

&lt;211&gt; 9323

&lt;212&gt; DNA

&lt;213&gt; Haemophilus influenzae

&lt;400&gt; 6

```

cgccacttca attttggatt gttgaaattc aactaaccaa aaagtgcggt taaaatctgt 60
ggagaaaaata ggttgtagtg aagaacgagg taattgttca aaaggataaa gctctcttaa 120
ttgggcattg gttggcggtt ctttttcggt taatagtaaa ttatattctg gacgactatg 180
caatccacca acaactttac cgttgggttt aagcggttaat gtaagttctt gctcttcttg 240
gcgaatacgt aatcccattt tttgttttagc aagaaaaatga tcgggataat cataataggt 300
gttgcccaaa aataaatttt gatgttctaa aatcataaat tttgcaagat attgtggcaa 360
ttcaatacct atttgtggcg aaatcgccaa ttttaattca atttcttgta gcataatatt 420
tcccactcaa atcaactggt taaatataca agataataaa aataaatcaa gatttttggtg 480
atgacaaaaca acaattacaa cacctttttt gcagtctata tgcaaatatt ttaaaaaaat 540
agtataaatc cgccatataa aatgggtataa tctttcatct ttcattcttc atctttcatc 600
tttcatcttt catctttcat ctttcatctt tcatctttca tctttcatct ttcattcttc 660
atctttcatc tttcatcttt cacatgaaat gatgaaccga gggaaggag ggaggggcaa 720
gaatgaagag ggagctgaac gaacgcaaat gataaagtaa ttaattgtt caactaacct 780
taggagaaaa tatgaacaag atatatcgtc tcaaattcag caaacgcctg aatgctttgg 840
ttgctgtgtc tgaattggca cggggttggt cacttagcgt taaagccact tccgctatg ttactatctt 960
ctgctgcgat gaaagtgcgt cacttagcgt taaagccact tccgctatg ttactatctt 960
taggtgtaac atctattcca caatctgttt tagcaagcgg caatttaaca tcgacaaaaa 1020
tgaaatggtg cagttttttac aagaaaacaa gtaataaaac cattatccgc aacagtgttg 1080
acgctatcat taattgaaa caatttaaca tcgacaaaaa tgaaatggtg cagttttttac 1140
aagaaaacaa caactccgcc gtattcaacc gtgttacatc taaccaaata tcccaattaa 1200
aagggtatgt agattctaac ggacaagtct ttttaataca ccaaatgggt atcacaatag 1260
gtaaagacgc aattattaac actaatggct ttacggcttc tacgctagac atttctaacg 1320
aaaacatcaa ggcgcgtaat ttcaccttcg agcaaaccaa agataaagcg ctgcgtgaaa 1380
ttgtgaatca cggtttaatt actgtcggtg aagacggcag tgtaaatctt attggtggca 1440
aagtgaaaaa cgagggtgtg attagcgtaa atggtggcag catttcttta ctgcgaggcg 1500
aaaaaatcac catcagcgat ataataaacc caaccattac ttacagcatt gccgcgcctg 1560
aaaatgaagc ggtcaatctg ggcgatattt ttgccaaagg cggtaacatt aatgtccgtg 1620
ctgccactat tcgaaaccaa ggtaaaactt ctgctgattc tgtaagcaaa gataaaagcg 1680
gcaatattgt tctttccgcc aaagagggtg aagcggaaat tggcggtgta atttccgctc 1740
aaaatcagca agctaaaggc ggcaagctga tgataaagtc cgataaagtc acattaaaaa 1800
caggtgcagt tatcgacctt tcaggtaaag aagggggaga aacttacctt ggcggtgacg 1860
agcgcgcgca aggtaaaaac ggcaattcaat tagcaaagaa aacctcttta gaaaaaggct 1920
caaccatcaa tgtatcaggc aaagaaaaag gcggacgcgc tattgtgtgg ggcgatattg 1980
cgtaattga cggcaatatt aacgctcaag gtagtgggtg tatcgctaaa accggtggtt 2040
ttgtggagac atcggggcgt tatttatcca ttgacagcaa tgcaattgtt aaaacaaaag 2100
agtgtgtgct agaccctgat gatgtaacaa ttgaagccga agacccccct cgcaataata 2160
ccggtataaa tgatgaattc ccaacaggca ccggtgaagc aagcgaccct aaaaaaata 2220
gcgaactcaa aacaacgcta accaatacaa ctatttcaaa ttatctgaaa aacgcctgga 2280

```

caatgaatat	aacggcatca	agaaaactta	ccgttaatag	ctcaatcaac	atcgggaagca	2340
actcccactt	aattctccat	agtaaagggtc	agcgtggcgg	aggcgttcag	attgatggag	2400
atattacttc	taaaggcgga	aatttaacca	tttattctgg	cggtatgggtt	gatgttcata	2460
aaaatattac	gcttgatcag	ggttttttta	atattaccgc	cgcttccgta	gcttttgaag	2520
gtggaaataa	caaagcacgc	gacgcggcaa	atgctaaaat	tgctgcccag	ggcactgtaa	2580
ccattacagg	agagggaaaa	gatttcaggg	ctaacaacgt	atctttaaac	ggaacgggta	2640
aaggctcgaa	tatcatttca	tcagtgaata	atttaaccca	caatcttagt	ggcacaatta	2700
acatatctgg	gaatataaca	attaacccaa	ctacgagaaa	gaacacctcg	tattggcaaa	2760
ccagccatga	ttcgcaactg	aacgtcagtg	ctcttaatct	agagacaggc	gcaaatttta	2820
cctttattaa	atacatttca	agcaatagca	aaggcttaac	aacacagtat	agaagctctg	2880
caggggtgaa	ttttaacggc	gtaaatggca	acatgtcatt	caatctcaaa	gaaggagcga	2940
aagttaattt	caaattaaaa	ccaaacgaga	acatgaacac	aagcaaactt	ttaccaattc	3000
ggtttttagc	caatatcaca	gccactgggtg	ggggctctgt	tttttttgat	atatatgcca	3060
accattctgg	cagagggggt	gagttaaaaa	tgagtgaat	taatatctct	aacggcgcta	3120
attttacctt	aaattcccat	gttcgcggcg	atgacgcttt	taaaatcaac	aaagacttaa	3180
ccataaatgc	aaccaattca	aatttcagcc	tcagacagac	gaaagatgat	ttttatgacg	3240
ggtacgcacg	caatgccatc	aattcaacct	acaacatatc	cattctgggc	ggtaatgtca	3300
cccttggtgg	acaaaactca	agcagcagca	ttacggggaa	tattactatc	gagaaagcag	3360
caaatgttac	gctagaagcc	aataacgccc	ctaatcagca	aaacataagg	gatagagtta	3420
taaaacttgg	cagcttgctc	gttaatggga	gtttaagttt	aactggcgaa	aatgcagata	3480
ttaaaggcaa	tctcactatt	tcagaaaagcg	ccacttttaa	aggaaagact	agagataccc	3540
taaatatcac	cggcaatttt	accaataatg	gcactgccga	aattaatata	acacaaggag	3600
tggtaaaaact	tggcaatgtt	accaatgatg	gtgattttaa	cattaccact	cacgctaacc	3660
gcaaccaaag	aagcatcatc	ggcggagata	taatcaacaa	aaaaggaagc	ttaaatatta	3720
cagacagtaa	taatgatgct	gaaatccaaa	ttggcggcaa	tatctcgcaa	aaagaaggca	3780
acctcacgat	ttcttccgat	aaaattaata	tcaccaaaca	gataacaatc	aaaaagggta	3840
ttgatggaga	ggactctagt	tcagatgcga	caagtaatgc	caacctaaact	attaaaacca	3900
aagaattgaa	attgacagaa	gacctaaagta	tttcagggtt	caataaagca	gagattacag	3960
ccaaagatgg	tagagattta	actattggca	acagtaatga	cggtaacagc	ggtagccgaag	4020
ccaaaacagt	aacttttaac	aatgttaaag	attcaaaaaat	ctctgctgac	ggtcacaatg	4080
tgcaactaaa	tagcaaagtg	aaaacatcta	gcagcaatgg	cggacgtgaa	agcaatagcg	4140
acaacgatac	cggcttaact	attactgcaa	aaaatgtaga	agtaaacaaa	gatattactt	4200
ctctcaaaac	agtaaataatc	accgcgtcgg	aaaagggttac	caccacagca	ggctcgacca	4260
ttaacgcaac	aatgggcaaa	gcaagtatta	caaccaaacc	aggtgatatac	agcggtagca	4320
tttcgggtaa	cacggtaagt	gttagcgcg	ctgggtgattt	aaccactaaa	tccgggtcaa	4380
aaattgaagc	gaaatcgggt	gaggctaattg	taacaagtgc	aacagggtaca	attggcggta	4440
caatttcggg	taatacggta	aatgttacgg	caaacgctgg	cgatttaaca	gttgggaatg	4500
gcgagaaaat	taatgcgaca	gaaggagctg	caaccttaac	cgcaacaggg	aataccttga	4560
ctactgaagc	cggttctagc	atcacttcaa	ctaagggtca	ggtagacctc	ttggctcaga	4620
atggttagct	gcaggaagc	attaatgctg	attaatgtgac	acaggaact	acaggcact	4680
taaccaccgt	ggcaggctcg	gatattaaag	caaccagcgg	caccttggtt	attaacgcaa	4740
aagatgctaa	gctaaatgg	gatgcacag	gtgatagtac	agaagtgaat	gcagtcaacg	4800
actggggatt	tggtagtggt	actgcggcaa	cctcaagcag	tgtgaatatac	actggggatt	4860
taaacacagt	aatgggtta	aatatcattt	cgaaagatgg	tagaaacact	gtgcgcttaa	4920
gaggcaagga	aattgaggtg	aatatatatc	agccagggtg	agcaagtgtta	gaagaagtaa	4980
ttgaagcgaa	acgcgtcctt	gaaaaagtaa	aagattttatc	tgatgaagaa	agagaaacat	5040
tagctaaact	tgggtgaagt	gctgtacgtt	ttgttgagcc	aaataatata	attacagtca	5100
atacacaaaa	tgaattttaca	accagaccgt	caagtcaagt	gataatttct	gaaggtaagg	5160
cgtgtttctc	aagtggtaat	ggcgcacgag	tatgtaccaa	tggtgctgac	gatggacagc	5220
cgtagtacgt	aattgacaag	gtagatttca	tcttgcaatg	aagtcatttt	attttcgtat	5280
tatttactgt	gtgggttaaa	gttcagtagc	ggctttaccc	atcttgtaaa	aaattacgga	5340
gaatacaata	aggtattttt	aacagggttat	tattatgaaa	aatataaaaa	gcagattaaa	5400
actcagtgc	atatcagtat	tgcttgccct	ggcttcttca	tcattgtatg	cagaagaagc	5460
gttttttagta	aaaggctttc	agttatctgg	tgacttgtaa	actttaagtg	aagacgcccc	5520
actgtctgta	gcaaaatctt	tatctaaata	ccagggtcgt	caaactttta	caaacctaaa	5580
aacagcacag	cttgaattac	aggctgtgct	agataagatt	gagccaaata	aatttgatgt	5640
gatattgccg	caacaaacca	ttacggatgg	caatatcatg	tttgagctag	tctcgaaatc	5700
agccgcagaa	agccaagttt	tttataaggc	gagccagggt	tatagtgaag	aaaatatcgc	5760
tcgtagcctg	ccatctttga	aacaaggaaa	agtgtatgaa	gatggctcgtc	agtggttcga	5820

tttgcgtgaa	tttaatatgg	caaaagaaaa	cccgccttaag	gttaccctg	tacattacga	5880
actaaaccct	aaaaacaaaa	cctctaattt	gataattg	ggcttctcgc	cttttggttaa	5940
aacgcgtagc	tttatttctt	atgataattt	cggcgcgaga	gagtttaact	accaacgtgt	6000
aagcttgggt	tttggttaatg	ccaattttaac	tggctcatgat	gatgtgttaa	ttataccagt	6060
atgagttatg	ctgatttctaa	tgatatcgac	ggcttaccaa	gtgcgattaa	tcgtaaatta	6120
tcaaaaggct	aatctatctc	tgcaaatctg	aaatggagtt	attatctccc	aacatttaac	6180
cttggcatgg	aagaccaatt	taaaattaat	ttaggctaca	actaccgcca	tattaatcaa	6240
acctccgct	taaatcgctt	gggtgaaacg	aagaaaaaat	ttgcagtatc	aggcgtaagt	6300
gcaggcattg	atggacatat	ccaatttacc	cctaaaacaa	tctttaatat	tgattttaact	6360
catcattatt	acgcgagtaa	attaccaggc	tcttttggaa	tggagcgcat	tggcgaaaca	6420
tttaatcgca	gctatcacat	tagcacagcc	agtttaggg	tgagtcaaga	gtttgctcaa	6480
ggttggcatt	ttagcagtc	attatcaggt	caatttactc	tacaagatat	tagcagtata	6540
gatttattct	ctgtaacagg	tacttatggc	gtcagaggct	ttaaatacgg	cgggtgcaagt	6600
gggtgagcgc	gtcttgtatg	gcgtaatgaa	ttaagtatgc	caaaatacac	ccgcttccaa	6660
atcagccctt	atgcgtttta	tgatgcaggt	cagttccggt	ataatagcga	aaatgctaaa	6720
acttacggcg	aagatatgca	cacgggtatcc	tctgcgggtt	taggcattaa	aacctctcct	6780
acacaaaact	taagcctaga	tgcttttgtt	gctcgtcgct	ttgcaaatgc	caatagtac	6840
aatttgaaatg	gcaacaaaaa	acgcacaagc	tcacctacaa	ccttctgggg	gagattaa	6900
ttcagtttct	aacctgaaa	tttaatcaac	tggtaaagcgt	tccgcctacc	agtttataac	6960
tatatgcttt	acccgccaat	ttacagtcta	taggcaaccc	tgtttttacc	cttatatact	7020
aaataaaca	gctaagctga	gctaagcaaa	ccaagcaaac	tcaagcaagc	caagtaatac	7080
taaaaaaaca	atttatatga	taaactaaag	tatactccat	gccatggcga	tacaagggat	7140
ttataaatat	gacaaaagaa	aatttgcaaa	acgctcctca	agatgcgacc	gctttacttg	7200
cggaattaag	caacaatcaa	actcccctgc	gaatatttaa	acaaccacgc	aagcccagcc	7260
tattacgctt	ggaacaacat	atcgcaaaaa	aagattatga	gtttgcttgt	cgtgaattaa	7320
tggtgattct	ggaaaaaatg	gacgctaatt	ttggaggcgt	tcacgatatt	gaatttgacg	7380
caccgcgtca	gctggcatat	ctacccgaaa	aattactaat	ttattttgcc	actcgtctcg	7440
ctaattgcaat	tacaacactc	ttttccgacc	ccgaattggc	aatttctgaa	gaaggggctg	7500
taaagatgat	tagcctgcaa	cgctgggtga	cgctgatttt	tgccctctcc	ccctacgtta	7560
acgcagacca	tattctcaat	aaatataata	tcaacccaga	ttccgaagg	ggctttcatt	7620
tagcaacaga	caactcttct	attgctaaat	tctgtatttt	ttacttacc	gaatccaatg	7680
tcaatatgag	tttagatg	ttatgggcag	ggaatcaaca	actttgtgct	tcatttgtgt	7740
ttgcgttgca	gtcttcacgt	tttattggta	ccgcatctgc	gtttcataaa	agagcgggtg	7800
ttttacagt	gtttcctaaa	aaactcgccg	aaattgctaa	tttagatgaa	ttgcctgcaa	7860
atatacttca	tgatgtatat	atgcactgca	gttatgattt	agcaaaaaac	aagcacgatg	7920
tttagcgtcc	attaacgaa	cttgccgca	agcatatcct	cacgcaagga	tggcaagacc	7980
gctaccttta	caccttaggt	aaaaaggacg	gcaaacctgt	gatgatggta	ctgcttgaac	8040
attttaattc	gggacattcg	atttatcgta	cacattcaac	ttcaatgatt	gctgctcgag	8100
aaaaattcta	tttagtcggc	ttaggccatg	agggcggtga	taaaataggt	cgagaagtgt	8160
ttgacgagtt	ctttgaaatc	agtagcaata	atataatgga	gagactgttt	tttatccgta	8220
aacagtgcca	aactttccaa	cccgagtggt	tctatatgcc	aagcattggc	atggatatta	8280
ccacgatttt	tgtgagcaac	actcggtgtg	ccctatttca	agctgtagcc	ctgggtcatc	8340
ctgccactac	gcattctgaa	tttattgatt	atgtcatcgt	agaagatgat	tatgtgggca	8400
gtgaagattg	tttcagcgaa	acccttttac	gcttacccaa	agatgcccta	ccttatgtac	8460
cttctgcact	cgccccacaa	aaagtggatt	atgtactcag	ggaaaaccct	gaagtagtca	8520
atatacggtat	tgccgctacc	acaatgaaat	taaacctga	atttttgcta	acattgcaag	8580
aatcagaga	taaagctaaa	gtcaaaatac	attttcattt	cgcaacttga	caatcaacag	8640
gcttgacaca	cccttatgtc	aaatggttta	tcgaaagcta	tttaggtgac	gatgccactg	8700
cacatcccca	cgcaccttat	cacgattatc	tggcaatatt	gcgtgattgc	gatattgtac	8760
taaatccggt	tcctttcggt	aataactaacg	gcataattga	tatgggtaca	ttaggtttag	8820
ttggtgtatg	caaaacgggg	gatgaagtac	atgaacatat	tgatgaagg	ctgttttaac	8880
gcttaggact	accagaatgg	ctgatagccg	acacacgaga	aacatatatt	gaatgtgctt	8940
tgcttctagc	agaaaacat	caagaacgcc	ttgaactccg	tcgttacatc	atagaaaaca	9000
acggcttaca	aaagcttttt	acaggcgacc	ctcgtccatt	gggcaaaata	ctgcttaaga	9060
aaacaaatga	atggaagcgg	aagcacttga	gtaaaaata	acggtttttt	aaagtaaaag	9120
tgcggttaat	tttcaaagcg	ttttaaaaaac	ctctcaaaaa	tcaaccgcac	ttttatcttt	9180
ataacgatcc	cgcacgctga	cagtttatca	gcctcccgcc	ataaaaactcc	gcctttcatg	9240
gcggagattt	tagccaaaaac	tggcagaaat	taagggttaa	aatcaccaaa	ttgcaccaca	9300
aatcaccaa	taccacaaa	aaa				9323

<210> 7  
 <211> 4794  
 <212> DNA  
 <213> Haemophilus influenzae

<400> 7  
 atgaacaaga tatatcgtct caaatccagc aaacgcctga atgcttttgg tgcgtgtgtct 60  
 gaattgacac ggggttgtga ccattccaca gaaaaaggca gtgaaaacc tgttcgtacg 120  
 aaagtacgcc acttggcggt aaagccactt tccgctatat tgctatcttt gggcatggca 180  
 tccattccgc aatctgtttt agcgcggtt tacagggaa tgagcgtcgt acacggtaca 240  
 gcaaccatgc aagtagacgg caataaaacc actatccgta atagcgtcaa tgctatcatc 300  
 aattggaac aatttaacat tgacaaaat gaaatgggtgc agtttttaca agaaagcagc 360  
 aactctgccc ttttcaaccg tgttacatct gacaaaatct cccaattaaa agggatttta 420  
 gattctaacc gacaagtctt tttaatcaac ccaaatggta tcacaatagg taaagacgca 480  
 attattaaca ctaatggctt tactgcttct acgctagaca tttctaacga aaacatcaag 540  
 gcgcgtaatt tcacccttga gcaaaccag gataaagcac tcgctgaaat cgtgaatcac 600  
 gggttaatta cgttgggtaa agacggtagc gtaaacctta ttgggtggca agtgaaaaac 660  
 gagggcgtga tatcgtaaaa tggcggtagt atttctttac ttgcagggca aaaaatcacc 720  
 atcagcgata taataaatcc aaccatcact tacagcattg ctgcacctga aaacgaagcg 780  
 atcaatctgg gcgatatttt tgccaaaggt ggtaacatta atgtccgcgc tgccactatt 840  
 cgcaataaag gtaaaccttt tgccgactct gtaagcaaag ataaaagtgg taacatttgt 900  
 ctctctgcca aagaagggtga agcggaaatt ggcggtgtaa tttccgctca aaatcagcaa 960  
 gccaaagggt gtaagttgat gattacaggc gataaagtta cattgaaaac ggggtgcagt 1020  
 atcgaccttt cgggtaaaaga agggggagaa acttatcttg gcggtgacga gcgtggcgaa 1080  
 ggtaaaaacg gcattcaatt agcaaagaaa accacttttag aaaaaggctc aacaattaat 1140  
 gtgtcaggta aagaaaaagg tgggcgcgct attgtatggg gcgatattgc gtttaattgac 1200  
 ggcaatatta atgccccagg taaagatctg gctaaaactg gtgggtttgt ggagacgtcg 1260  
 gggcattact tatccattga tgataacgca attgttaaaa caaaaagaatg gctactagac 1320  
 ccagagaatg tgactattga agctccttcc gcttctcgcg tcgagctggg tgccgatagg 1380  
 aattccact cggcagaggt gataaaagt accctaaaaa aaaataaacac ctcttgaca 1440  
 aactaacca atacaacat ttcaaatctt ctgaaaagt cccacgtggg gaacataacg 1500  
 gcaaggagaa aacttaccgt taatagctct atcagtatag aaagaggctc ccacttaatt 1560  
 ctccacagtg aaggtcaggg cggcgaaggt gttcagattg ataaagatat tacttctgaa 1620  
 ggccgaaatt taaccattta ttctggcgga tgggttgatg ttcataaaaa tattacgctt 1680  
 ggtagcggct ttttaaacat cacaactaaa gaaggagata tcgccttoga agacaagtct 1740  
 ggacggaaca acctaaccat tacagcccaa gggaccatca cctcaggtaa tagtaacggc 1800  
 ttttagattta acaacgtctc tctaaacagc cttggcgga agctgagctt tactgacagc 1860  
 agagaggaca gaggtagaag aactaagggt aatatctcaa acaaatttga cggaaactgt 1920  
 aacatttccg gaactgtaga tatctcaatg aaagcaccca aagtcagctg gttttacaga 1980  
 gacaaaggac gcacctactg gaacgtaacc actttaaatg ttacctcggg tagtaaat 2040  
 aacctctcca ttgacagcac aggaagtggc tcaacaggct caagcatacg caatgcagaa 2100  
 ttaaattggca taacatttaa taaagccact tttaatatcg cacaaggctc aacagctaac 2160  
 ttttagcatca aggcatacat aatgcccttt aagagtaacg ctaactacgc attatttaat 2220  
 gaagatattt cagtctcagg ggggggtagc cttaatttca aacttaacgc ctcatctagc 2280  
 aacatacaaa cccctggcgt aattataaaa tctcaaaact ttaatgtctc aggagggtca 2340  
 actttaaatc tcaaggctga aggttcaaca gaaaccgctt tttcaataga aaatgattta 2400  
 aacttaaacg ccaccgggtg caatataaca atcagacaag tcgagggtac cgattcacgc 2460  
 gtcaacaaag gtgtcgcagc caaaaaaac ataactttta aagggggtta tatcaccttc 2520  
 ggctctcaaa aggccacaac agaaatcaaa ggcaatgtta ccatcaataa aaacactaac 2580  
 gctactcttt gtgggtcgaa ttttgcgaa acaaatcgc ctttaaatat agcaggaaat 2640  
 gttattaata atggcaacct taccactgcc ggctccatta tcaatatagc cggaaatctt 2700  
 actgtttcaa aaggcgctaa cttcaagct ataacaaatt acacttttaa tgtagccggc 2760  
 tcatttgaca acaatggcgc ttcaaacatt tccattgcca gaggaggggc taaattttaa 2820  
 gatatacaata acaccagtag cttaaatatt accaccaact ctgataccac ttaccgcacc 2880  
 attataaaaag gcaatatatc caacaaatca ggtgatttga atattattga taaaaaaagc 2940  
 gacgctgaaa tccaaattgg cggcaatatc tcacaaaaag aaggcaatct cacaatttct 3000  
 tctgataaag taaatattac caatcagata acaatcaaag caggcgttga aggggggcgt 3060

```

tctgattcaa gtgaggcaga aaatgctaac ctaactattc aaaccaaaga gttaaaattg 3120
gcaggagacc taaatatttc aggccttaat aaagcagaaa ttacagctaa aaatggcagt 3180
gatttaacta ttggcaatgc tagcgggtgt aatgctgatg ctaaaaaagt gacttttgac 3240
aagggttaaag attcaaaaat ctgcactgac ggtcacaatg taacactaaa tagcgaagtg 3300
aaaacgtcta atggtagtag caatgctggg aatgataaca gcaccggtt aaccatttcc 3360
gcaaaagatg taacggtaaa caataacggt acctcccaca agacaataaa tatctctgcc 3420
gcagcaggaa atgtaacaac caaagaaggc acaactatca atgcaaccac aggcagcgtg 3480
gaagtaactg ctcaaaatgg tacaattaaa ggcaacatta cctcgcaaaa tgtaacagtg 3540
acagcaacag aaaatcttgt taccacagag aatgctgtca ttaatgcaac cagcggcaca 3600
gtaaacatta gtacaaaaac aggggatatt aaaggtggaa ttgaatcaac ttccggtaat 3660
gtaaatatta cagcgagcgg caatacactt aaggttaagta atatcactgg tcaagatgta 3720
acagtaacag cggatgcagg agccttgaca actacagcag gctcaaccat tagtgcgaca 3780
acaggcaatg caaatattac aaccaaaca ggtgatatca acggtaaagt tgaatccagc 3840
tccggctctg taacacttgt tgcaactgga gcaactcttg ctgtaggtaa tatttcaggt 3900
aacactgtta ctattactgc ggatagcggg aaattaacct ccacagtagg ttctacaatt 3960
aatgggacta atagtgtaac cacctcaagg caatcaggcg atattgaagg tacaatttct 4020
ggtaatacag taaatgttac agcaagcact ggtgatttaa ctattggaaa tagtgcaaaa 4080
ggtgaagcga aaaatggagc tgcaacctta actgctgaat caggcaaatt aaccacccaa 4140
acaggctcta gcattacctc aagcaatggg cagacaactc ttacagccaa ggatagcagt 4200
atcgaggaa acattaatgc tgctaattgt acgttaaata ccacaggcac tttaactact 4260
acaggggatt caaagattaa cgcaaccagt ggtaccttaa caatcaatgc aaaagatgcc 4320
aaattagatg tgctgtcatc aggtgaccgc acagtagtaa atgcaactaa cgcaagtggc 4380
tctggtaacg tgactgcgaa aacctcaagg agcgtgaata tcaccgggga tttaaacaca 4440
ataaatgggt taaatatcat ttcggaat ggtagaaaca ctgtgcgctt aagaggcaag 4500
gaaattgatg tgaaatatat ccaaccaggt gtagcaagcg tagaagaggt aattgaagcg 4560
aaacgcgtcc ttgagaaggt aaaagattta tctgatgaag aaagagaaac actagccaaa 4620
cttgggtgta gtgctgtacg tttcgttgag ccaataatg ccattacggg taatacacia 4680
aacgagttta caaccaaacc atcaagtcaa gtgacaattt ctgaaggtaa ggcgtgtttc 4740
tcaagtggta atggcgacag agtatgtacc aatgttgctg acgatggaca gcag 4794

```

&lt;210&gt; 8

&lt;211&gt; 4803

&lt;212&gt; DNA

&lt;213&gt; Haemophilus influenzae

&lt;400&gt; 8

```

atgaacaaga tatatcgtct caaattcagc aaacgcctga atgctttggg tgctgtgtct 60
gaattgacac ggggttgatg ccattccaca gaaaaaggca gtgaaaaacc tgttcgtacg 120
aaagtaacgc acttgccgtt aaagccactt tccgctatat tgctatcttt gggcatggca 180
tccattccgc aatctgtttt agcgagcggg taacagggaa tgagcgtcgt acacgggtaca 240
gcaaacatgc aagtagacgg caataaaaacc actatccgta atagcgtcaa tgctatcatc 300
aattggaaac aatttaacat tgacaaaat gaaatgggtg agtttttaca agaaagcagc 360
aactctgccg ttttcaaccg tgttacatct gaccaaatct cccaattaaa agggatttta 420
gattctaacg gacaagtctt tttaatcaac ccaaatggta tcacaatagg taaagacgca 480
attattaaca ctaatggctt tactgcttct acgctagaca tttctaacga aaacatcaag 540
gcgcgtaatt tcacccttga gcaaaccaag gataaagcac tcgctgaaat cgtgaatcac 600
ggtttaatta ccgttggtaa agacggtagc gtaaacctta ttggtggcaa agtgaaaaac 660
gagggcggtg ttagcgtaaa tggcggtagt atttctttac ttgcagggca aaaaatcacc 720
atcagcgata taataaatcc aaccatcact tacagcattg ctgcacctga aaacgaagcg 780
atcaatctgg gcgatatatt tgccaaaagg ggtaacatta atgtccgcgc tgccactatt 840
cgcaataaag gtaaaccttc tgccgactct gtaagcaaaag ataaaagtgg taacattgtt 900
ctctctgcca aagaagggtg agcggaaatt ggcgggtgaa tttccgctca aaatcagcaa 960
gccaaagggt gtaagttgat gattacaggt gataaagtca cattaataaac aggtgcagtt 1020
atcgaccttt caggtaaaga agggggagag acttatcttg gcggtgatga gcgtggcgaa 1080
ggtaaaaatg gtattcaatt agcgaagaaa acctctttag aaaaaggctc gacaattaat 1140
gtatcaggca aagaaaaagg cgggcgcgct attgtatggg gcgatattgc attaatatt 1200
ggtaacatta atgctcaagg tagcgatatt gctaaaactg gcggctttgt ggaaacatca 1260
ggacatgact tatccattgg tgatgatgtg attgttgacg ctaaagagtg gttattagac 1320

```

ccagatgatg tgtccattga aactcttaca tctggacgca ataataccgg cgaaaaccaa 1380  
 ggatatacaa caggagatgg gactaaagag tcacctaaag gtaatagtat ttctaaacct 1440  
 acattaacaa actcaactct tgagcaaatc ctaagaagag gttcttatgt taatatcact 1500  
 gctaataata gaatttatgt taatagctcc atcaacttat ctaatggcag ttttaacactt 1560  
 cacactaaac gagatggagt taaaattaac ggtgatatta cctcaaacga aaatggtaat 1620  
 ttaaccatta aagcaggctc ttgggttgat gttcataaaa acatcacgct tggtagcggg 1680  
 tttttgaata ttgtcgctgg ggattctgta gcttttgaga gagaggcgga taaagcacgt 1740  
 aacgcaacag atgctcaaat taccgcacaa gggacgataa ccgtaataa agatgataaa 1800  
 caatttagat tcaataatgt atctattaac gggacgggca aggggtttaa gtttattgca 1860  
 aatcaaaata atttcaactca taaatttgat ggcgaaatta acatatctgg aatagtaaca 1920  
 attaaccaaa ccacgaaaaa agatgttaa tactggatg catcaaaaaga ctcttactgg 1980  
 aatgtttctt ctcttacttt gaatacgggtg caaaaattta cttttataaa attcgttgat 2040  
 agcggctcaa attcccaaga ttgagggtca tcacgtagaa gttttgcagg cgtacatttt 2100  
 aacggcatcg gaggcaaaac aaacttcaac atcggagcta acgcaaaaagc cttattttaa 2160  
 ttaaaaccaa acgcccgtac agaccacaaa aaagaattac ctattacttt taacgccaac 2220  
 attacagcta ccggtaacag tgatagctct gtgatgtttg acatacacgc caatcttacc 2280  
 tctagagctg ccggcataaa catggattca attaacatta ccggcgggct tgacttttcc 2340  
 ataacatccc ataactgcaa tagtaatgct tttgaaatca aaaaagactt aactataaat 2400  
 gcaactggct cgaatttttag tcttaagcaa acgaaagatt ctttttataa tgaatacagc 2460  
 aaacacgcca ttaactcaag tcataatcta accattcttg gcggcaatgt cactctaggt 2520  
 ggggaaaatt caagcagtag cattacgggc aatatcaata tcaccaataa agcaaatgtt 2580  
 acattacaag ctgacaccag caacagcaac acaggcttga agaaaagaac tctaactctt 2640  
 ggcaatatat ctgttgaggg gaatttaagc ctaactggtg caaatgcaaa cattgtcggc 2700  
 aatctttcta ttgcagaaga ttccacattt aaaggagaag ccagtgacaa cctaaacatc 2760  
 accggcacct ttaccaacaa cggtagcgcc aacattaata taaaacaagg agtggtaaaa 2820  
 ctccaaggcg atattatcaa taaagggtgt taaatatca ctactaacgc ctcaggcact 2880  
 caaaaaacca ttattaacgg aaatataact aacgaaaaag gcgacttaa catcaagaat 2940  
 attaaagccg acgcccgaat ccaaattggc ggcaatatct caaaaaaga aggcattctc 3000  
 acaatttctt ctgataaagt aaatattacc aatcagataa caatcaaagc aggcgttgaa 3060  
 gggggcgctt ctgattcaag tgaggcagaa aatgctaacc taactattca aaccaaagag 3120  
 ttaaaatttg caggagacct aaatatttca ggctttaata aagcagaaat tacagctaaa 3180  
 aatggcagtg atttaactat tggcaatgct agcgggtgta atgctgatgc taaaaaagt 3240  
 acttttgaca aggttaaaga ttcaaaaatc tcgactgacg gtcacaatgt aacactaaat 3300  
 agcgaagtga aaacgtctaa tggtagtagc aatgctggtg atgataacag caccgggtta 3360  
 accatttccg caaaagatgt aacggtaaac aataacgtta cctcccacaa gacaataaat 3420  
 atctctgccc cagcaggaaa tgtaacaacc aaagaaggca caactatcaa tgcaaccaca 3480  
 ggcagcgtgg aagtaactgc tcaaaatggg acaattaaag gcaacattac ctgcacaaat 3540  
 gtaacagtga cagcaacaga aaatcttggt accacagaga atgctgtcat taatgcaacc 3600  
 agcggcacag taaacattag tacaaaaaca ggggatatta aagggtggaat tgaatcaact 3660  
 tccggtaattg taaatattac agcagcggc aatacactta aggttaagtaa tatcactggt 3720  
 caagatgtaa cagtaacagc ggatgcagga gccttgacaa ctacagcagg ctcaaccatt 3780  
 agtgcgacaa caggcaatgc aaatattaca accaaaacag gtgatataa cggtaaagt 3840  
 gaatccagct ccggctctgt aacacttggt gcaactggag caactcttgc tgtaggtaat 3900  
 atttcaggta aactgttac tattactgag gatagcggta aattaacctc cacagtaggt 3960  
 tctacaatta atgggactaa tagtgaacc acctcaagcc aatcaggcga tattgaagg 4020  
 acaatttctg gtaatacagt aaatgttaca gcaagcactg gtgatttaac tattggaaat 4080  
 agtgcaaaaag ttgaagcgaa aaatggagct gcaaccttaa ctgctgaatc aggcacaaat 4140  
 accacccaaa caggctctag cattacctca agcaatgggt agacaactct tacagccaag 4200  
 gatagcagta tcgcaggaaa cattaatgct gctaattgta cgttaaatac cacaggcact 4260  
 ttaactacta caggggattc aaagattaac gcaaccagtg gtaccttaac aatcaatgca 4320  
 aaagatgcc aattagatgg tgctgcatca ggtgaccgca cagtagtaaa tgcaactaac 4380  
 gcaagtggct ctggtaacgt gactgcgaaa acctcaagca gcgtgaatat caccggggat 4440  
 ttaaacacaa taaatgggtt aaatatcatt tcggaaaaat gtagaaacac tgtgcgctta 4500  
 agaggcaagg aaattgatgt gaaatatatc caaccaggtg tagcaagcgt agaagaggta 4560  
 attgaagcga aacgcgtcct tgagaaggta aaagatttat ctgatgaaga aagagaaaca 4620  
 ctagccaaac ttggtgtaag tgctgtacgt ttcgttgagc caaataatgc cattacgggt 4680  
 aatacacaaa acgagtttac aaccaaacca tcaagtcaag tgacaatttc tgaaggtaag 4740  
 gcgtgtttct caagtggtaa tggcgcacga gtatgtacca atgttgctga cgatggacag 4800  
 cag 4803

<210> 9  
 <211> 1599  
 <212> PRT  
 <213> Haemophilus influenzae

<400> 9  
 Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu  
     1                    5                    10                    15  
 Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys  
                     20                    25                    30  
 Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys  
                     35                    40                    45  
 Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln  
                     50                    55                    60  
 Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr  
                     65                    70                    75                    80  
 Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val  
                     85                    90                    95  
 Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
                     100                    105                    110  
 Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val  
                     115                    120                    125  
 Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
                     130                    135                    140  
 Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
                     145                    150                    155                    160  
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
                     165                    170                    175  
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys  
                     180                    185                    190  
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
                     195                    200                    205  
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
                     210                    215                    220  
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
                     225                    230                    235                    240  
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
                     245                    250                    255  
 Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
                     260                    265                    270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala  
 275 280 285  
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
 290 295 300  
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
 305 310 315 320  
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
 325 330 335  
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
 340 345 350  
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
 355 360 365  
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
 370 375 380  
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
 385 390 395 400  
 Gly Asn Ile Asn Ala Gln Gly Lys Asp Ile Ala Lys Thr Gly Gly Phe  
 405 410 415  
 Val Glu Thr Ser Gly His Tyr Leu Ser Ile Asp Asp Asn Ala Ile Val  
 420 425 430  
 Lys Thr Lys Glu Trp Leu Leu Asp Pro Glu Asn Val Thr Ile Glu Ala  
 435 440 445  
 Pro Ser Ala Ser Arg Val Glu Leu Gly Ala Asp Arg Asn Ser His Ser  
 450 455 460  
 Ala Glu Val Ile Lys Val Thr Leu Lys Lys Asn Asn Thr Ser Leu Thr  
 465 470 475 480  
 Thr Leu Thr Asn Thr Thr Ile Ser Asn Leu Leu Lys Ser Ala His Val  
 485 490 495  
 Val Asn Ile Thr Ala Arg Arg Lys Leu Thr Val Asn Ser Ser Ile Ser  
 500 505 510  
 Ile Glu Arg Gly Ser His Leu Ile Leu His Ser Glu Gly Gln Gly Gly  
 515 520 525  
 Gln Gly Val Gln Ile Asp Lys Asp Ile Thr Ser Glu Gly Gly Asn Leu  
 530 535 540  
 Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr Leu  
 545 550 555 560  
 Gly Ser Gly Phe Leu Asn Ile Thr Thr Lys Glu Gly Asp Ile Ala Phe  
 565 570 575  
 Glu Asp Lys Ser Gly Arg Asn Asn Leu Thr Ile Thr Ala Gln Gly Thr  
 580 585 590



Ile Thr Ser Gly Asn Ser Asn Gly Phe Arg Phe Asn Asn Val Ser Leu  
 595 600 605  
 Asn Ser Leu Gly Gly Lys Leu Ser Phe Thr Asp Ser Arg Glu Asp Arg  
 610 615 620  
 Gly Arg Arg Thr Lys Gly Asn Ile Ser Asn Lys Phe Asp Gly Thr Leu  
 625 630 635 640  
 Asn Ile Ser Gly Thr Val Asp Ile Ser Met Lys Ala Pro Lys Val Ser  
 645 650 655  
 Trp Phe Tyr Arg Asp Lys Gly Arg Thr Tyr Trp Asn Val Thr Thr Leu  
 660 665 670  
 Asn Val Thr Ser Gly Ser Lys Phe Asn Leu Ser Ile Asp Ser Thr Gly  
 675 680 685  
 Ser Gly Ser Thr Gly Pro Ser Ile Arg Asn Ala Glu Leu Asn Gly Ile  
 690 695 700  
 Thr Phe Asn Lys Ala Thr Phe Asn Ile Ala Gln Gly Ser Thr Ala Asn  
 705 710 715 720  
 Phe Ser Ile Lys Ala Ser Ile Met Pro Phe Lys Ser Asn Ala Asn Tyr  
 725 730 735  
 Ala Leu Phe Asn Glu Asp Ile Ser Val Ser Gly Gly Gly Ser Val Asn  
 740 745 750  
 Phe Lys Leu Asn Ala Ser Ser Ser Asn Ile Gln Thr Pro Gly Val Ile  
 755 760 765  
 Ile Lys Ser Gln Asn Phe Asn Val Ser Gly Gly Ser Thr Leu Asn Leu  
 770 775 780  
 Lys Ala Glu Gly Ser Thr Glu Thr Ala Phe Ser Ile Glu Asn Asp Leu  
 785 790 795 800  
 Asn Leu Asn Ala Thr Gly Gly Asn Ile Thr Ile Arg Gln Val Glu Gly  
 805 810 815  
 Thr Asp Ser Arg Val Asn Lys Gly Val Ala Ala Lys Lys Asn Ile Thr  
 820 825 830  
 Phe Lys Gly Gly Asn Ile Thr Phe Gly Ser Gln Lys Ala Thr Thr Glu  
 835 840 845  
 Ile Lys Gly Asn Val Thr Ile Asn Lys Asn Thr Asn Ala Thr Leu Arg  
 850 855 860  
 Gly Ala Asn Phe Ala Glu Asn Lys Ser Pro Leu Asn Ile Ala Gly Asn  
 865 870 875 880  
 Val Ile Asn Asn Gly Asn Leu Thr Thr Ala Gly Ser Ile Ile Asn Ile  
 885 890 895

Ala Gly Asn Leu Thr Val Ser Lys Gly Ala Asn Leu Gln Ala Ile Thr  
 900 905 910  
 Asn Tyr Thr Phe Asn Val Ala Gly Ser Phe Asp Asn Asn Gly Ala Ser  
 915 920 925  
 Asn Ile Ser Ile Ala Arg Gly Gly Ala Lys Phe Lys Asp Ile Asn Asn  
 930 935 940  
 Thr Ser Ser Leu Asn Ile Thr Thr Asn Ser Asp Thr Thr Tyr Arg Thr  
 945 950 955 960  
 Ile Ile Lys Gly Asn Ile Ser Asn Lys Ser Gly Asp Leu Asn Ile Ile  
 965 970 975  
 Asp Lys Lys Ser Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser Gln  
 980 985 990  
 Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr Asn  
 995 1000 1005  
 Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser Ser  
 1010 1015 1020  
 Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys Leu  
 1025 1030 1035 1040  
 Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala  
 1045 1050 1055  
 Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn Ala  
 1060 1065 1070  
 Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile Ser  
 1075 1080 1085  
 Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser Asn  
 1090 1095 1100  
 Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile Ser  
 1105 1110 1115 1120  
 Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr Ile  
 1125 1130 1135  
 Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr Thr  
 1140 1145 1150  
 Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly Thr  
 1155 1160 1165  
 Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr Glu  
 1170 1175 1180  
 Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly Thr  
 1185 1190 1195 1200  
 Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu Ser  
 1205 1210 1215

Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys Val  
 1220 1225 1230  
 Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly Ala  
 1235 1240 1245  
 Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn Ala  
 1250 1255 1260  
 Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser Ser  
 1265 1270 1275 1280  
 Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val Gly  
 1285 1290 1295  
 Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys Leu  
 1300 1305 1310  
 Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr Thr  
 1315 1320 1325  
 Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr Val  
 1330 1335 1340  
 Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala Lys  
 1345 1350 1355 1360  
 Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly Lys  
 1365 1370 1375  
 Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln Thr  
 1380 1385 1390  
 Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala Ala  
 1395 1400 1405  
 Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp Ser  
 1410 1415 1420  
 Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp Ala  
 1425 1430 1435 1440  
 Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala Thr  
 1445 1450 1455  
 Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser Val  
 1460 1465 1470  
 Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile Ser  
 1475 1480 1485  
 Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp Val  
 1490 1495 1500  
 Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu Ala  
 1505 1510 1515 1520

Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg Glu  
 1525 1530 1535

Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro Asn  
 1540 1545 1550

Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro Ser  
 1555 1560 1565

Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly Asn  
 1570 1575 1580

Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro  
 1585 1590 1595

<210> 10

<211> 1600

<212> PRT

<213> Haemophilus influenzae

<400> 10

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu  
 1 5 10 15

Val Ala Val Ser Glu Leu Thr Arg Gly Cys Asp His Ser Thr Glu Lys  
 20 25 30

Gly Ser Glu Lys Pro Val Arg Thr Lys Val Arg His Leu Ala Leu Lys  
 35 40 45

Pro Leu Ser Ala Ile Leu Leu Ser Leu Gly Met Ala Ser Ile Pro Gln  
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Ser Val Val His Gly Thr  
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Thr Ile Arg Asn Ser Val  
 85 90 95

Asn Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met  
 100 105 110

Glu Gln Phe Leu Gln Glu Ser Ser Asn Ser Ala Val Phe Asn Arg Val  
 115 120 125

Thr Ser Asp Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly  
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala  
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn  
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Leu Glu Gln Thr Lys Asp Lys  
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp  
 195 200 205  
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile  
 210 215 220  
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr  
 225 230 235 240  
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro  
 245 250 255  
 Glu Asn Glu Ala Ile Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn  
 260 265 270  
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Lys Gly Lys Leu Ser Ala  
 275 280 285  
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys  
 290 295 300  
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln  
 305 310 315 320  
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys  
 325 330 335  
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr  
 340 345 350  
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala  
 355 360 365  
 Lys Lys Thr Thr Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys  
 370 375 380  
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp  
 385 390 395 400  
 Gly Asn Ile Asn Ala Gln Gly Ser Asp Ile Ala Lys Thr Gly Gly Phe  
 405 410 415  
 Val Glu Thr Ser Gly His Asp Leu Ser Ile Gly Asp Asp Val Ile Val  
 420 425 430  
 Asp Ala Lys Glu Trp Leu Leu Asp Pro Asp Asp Val Ser Ile Glu Thr  
 435 440 445  
 Leu Thr Ser Gly Arg Asn Asn Thr Gly Glu Asn Gln Gly Tyr Thr Thr  
 450 455 460  
 Gly Asp Gly Thr Lys Glu Ser Pro Lys Gly Asn Ser Ile Ser Lys Pro  
 465 470 475 480  
 Thr Leu Thr Asn Ser Thr Leu Glu Gln Ile Leu Arg Arg Gly Ser Tyr  
 485 490 495  
 Val Asn Ile Thr Ala Asn Asn Arg Ile Tyr Val Asn Ser Ser Ile Asn  
 500 505 510

Leu Ser Asn Gly Ser Leu Thr Leu His Thr Lys Arg Asp Gly Val Lys  
 515 520 525  
 Ile Asn Gly Asp Ile Thr Ser Asn Glu Asn Gly Asn Leu Thr Ile Lys  
 530 535 540  
 Ala Gly Ser Trp Val Asp Val His Lys Asn Ile Thr Leu Gly Thr Gly  
 545 550 555 560  
 Phe Leu Asn Ile Val Ala Gly Asp Ser Val Ala Phe Glu Arg Glu Gly  
 565 570 575  
 Asp Lys Ala Arg Asn Ala Thr Asp Ala Gln Ile Thr Ala Gln Gly Thr  
 580 585 590  
 Ile Thr Val Asn Lys Asp Asp Lys Gln Phe Arg Phe Asn Asn Val Ser  
 595 600 605  
 Leu Asn Gly Thr Gly Lys Gly Leu Lys Phe Ile Ala Asn Gln Asn Asn  
 610 615 620  
 Phe Thr His Lys Phe Asp Gly Glu Ile Asn Ile Ser Gly Ile Val Thr  
 625 630 635 640  
 Ile Asn Gln Thr Thr Lys Lys Asp Val Lys Tyr Trp Asn Ala Ser Lys  
 645 650 655  
 Asp Ser Tyr Trp Asn Val Ser Ser Leu Thr Leu Asn Thr Val Gln Lys  
 660 665 670  
 Phe Thr Phe Ile Lys Phe Val Asp Ser Gly Ser Asn Gly Gln Asp Leu  
 675 680 685  
 Arg Ser Ser Arg Arg Ser Phe Ala Gly Val His Phe Asn Gly Ile Gly  
 690 695 700  
 Gly Lys Thr Asn Phe Asn Ile Gly Ala Asn Ala Lys Ala Leu Phe Lys  
 705 710 715 720  
 Leu Lys Pro Asn Ala Ala Thr Asp Pro Lys Lys Glu Leu Pro Ile Thr  
 725 730 735  
 Phe Asn Ala Asn Ile Thr Ala Thr Gly Asn Ser Asp Ser Ser Val Met  
 740 745 750  
 Phe Asp Ile His Ala Asn Leu Thr Ser Arg Ala Ala Gly Ile Asn Met  
 755 760 765  
 Asp Ser Ile Asn Ile Thr Gly Gly Leu Asp Phe Ser Ile Thr Ser His  
 770 775 780  
 Asn Arg Asn Ser Asn Ala Phe Glu Ile Lys Lys Asp Leu Thr Ile Asn  
 785 790 795 800  
 Ala Thr Gly Ser Asn Phe Ser Leu Lys Gln Thr Lys Asp Ser Phe Tyr  
 805 810 815

Asn Glu Tyr Ser Lys His Ala Ile Asn Ser Ser His Asn Leu Thr Ile  
 820 825 830  
 Leu Gly Gly Asn Val Thr Leu Gly Gly Glu Asn Ser Ser Ser Ser Ile  
 835 840 845  
 Thr Gly Asn Ile Asn Ile Thr Asn Lys Ala Asn Val Thr Leu Gln Ala  
 850 855 860  
 Asp Thr Ser Asn Ser Asn Thr Gly Leu Lys Lys Arg Thr Leu Thr Leu  
 865 870 875 880  
 Gly Asn Ile Ser Val Glu Gly Asn Leu Ser Leu Thr Gly Ala Asn Ala  
 885 890 895  
 Asn Ile Val Gly Asn Leu Ser Ile Ala Glu Asp Ser Thr Phe Lys Gly  
 900 905 910  
 Glu Ala Ser Asp Asn Leu Asn Ile Thr Gly Thr Phe Thr Asn Asn Gly  
 915 920 925  
 Thr Ala Asn Ile Asn Ile Lys Gly Val Val Lys Leu Gly Asp Ile Asn  
 930 935 940  
 Asn Lys Gly Gly Leu Asn Ile Thr Thr Asn Ala Ser Gly Thr Gln Lys  
 945 950 955 960  
 Thr Ile Ile Asn Gly Asn Ile Thr Asn Glu Lys Gly Asp Leu Asn Ile  
 965 970 975  
 Lys Asn Ile Lys Ala Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser  
 980 985 990  
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Val Asn Ile Thr  
 995 1000 1005  
 Asn Gln Ile Thr Ile Lys Ala Gly Val Glu Gly Gly Arg Ser Asp Ser  
 1010 1015 1020  
 Ser Glu Ala Glu Asn Ala Asn Leu Thr Ile Gln Thr Lys Glu Leu Lys  
 1025 1030 1035 1040  
 Leu Ala Gly Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr  
 1045 1050 1055  
 Ala Lys Asn Gly Ser Asp Leu Thr Ile Gly Asn Ala Ser Gly Gly Asn  
 1060 1065 1070  
 Ala Asp Ala Lys Lys Val Thr Phe Asp Lys Val Lys Asp Ser Lys Ile  
 1075 1080 1085  
 Ser Thr Asp Gly His Asn Val Thr Leu Asn Ser Glu Val Lys Thr Ser  
 1090 1095 1100  
 Asn Gly Ser Ser Asn Ala Gly Asn Asp Asn Ser Thr Gly Leu Thr Ile  
 1105 1110 1115 1120  
 Ser Ala Lys Asp Val Thr Val Asn Asn Asn Val Thr Ser His Lys Thr  
 1125 1130 1135

Ile Asn Ile Ser Ala Ala Ala Gly Asn Val Thr Thr Lys Glu Gly Thr  
 1140 1145 1150  
 Thr Ile Asn Ala Thr Thr Gly Ser Val Glu Val Thr Ala Gln Asn Gly  
 1155 1160 1165  
 Thr Ile Lys Gly Asn Ile Thr Ser Gln Asn Val Thr Val Thr Ala Thr  
 1170 1175 1180  
 Glu Asn Leu Val Thr Thr Glu Asn Ala Val Ile Asn Ala Thr Ser Gly  
 1185 1190 1195 1200  
 Thr Val Asn Ile Ser Thr Lys Thr Gly Asp Ile Lys Gly Gly Ile Glu  
 1205 1210 1215  
 Ser Thr Ser Gly Asn Val Asn Ile Thr Ala Ser Gly Asn Thr Leu Lys  
 1220 1225 1230  
 Val Ser Asn Ile Thr Gly Gln Asp Val Thr Val Thr Ala Asp Ala Gly  
 1235 1240 1245  
 Ala Leu Thr Thr Thr Ala Gly Ser Thr Ile Ser Ala Thr Thr Gly Asn  
 1250 1255 1260  
 Ala Asn Ile Thr Thr Lys Thr Gly Asp Ile Asn Gly Lys Val Glu Ser  
 1265 1270 1275 1280  
 Ser Ser Gly Ser Val Thr Leu Val Ala Thr Gly Ala Thr Leu Ala Val  
 1285 1290 1295  
 Gly Asn Ile Ser Gly Asn Thr Val Thr Ile Thr Ala Asp Ser Gly Lys  
 1300 1305 1310  
 Leu Thr Ser Thr Val Gly Ser Thr Ile Asn Gly Thr Asn Ser Val Thr  
 1315 1320 1325  
 Thr Ser Ser Gln Ser Gly Asp Ile Glu Gly Thr Ile Ser Gly Asn Thr  
 1330 1335 1340  
 Val Asn Val Thr Ala Ser Thr Gly Asp Leu Thr Ile Gly Asn Ser Ala  
 1345 1350 1355 1360  
 Lys Val Glu Ala Lys Asn Gly Ala Ala Thr Leu Thr Ala Glu Ser Gly  
 1365 1370 1375  
 Lys Leu Thr Thr Gln Thr Gly Ser Ser Ile Thr Ser Ser Asn Gly Gln  
 1380 1385 1390  
 Thr Thr Leu Thr Ala Lys Asp Ser Ser Ile Ala Gly Asn Ile Asn Ala  
 1395 1400 1405  
 Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Thr Gly Asp  
 1410 1415 1420  
 Ser Lys Ile Asn Ala Thr Ser Gly Thr Leu Thr Ile Asn Ala Lys Asp  
 1425 1430 1435 1440



Ala Lys Leu Asp Gly Ala Ala Ser Gly Asp Arg Thr Val Val Asn Ala  
1445 1450 1455

Thr Asn Ala Ser Gly Ser Gly Asn Val Thr Ala Lys Thr Ser Ser Ser  
1460 1465 1470

Val Asn Ile Thr Gly Asp Leu Asn Thr Ile Asn Gly Leu Asn Ile Ile  
1475 1480 1485

Ser Glu Asn Gly Arg Asn Thr Val Arg Leu Arg Gly Lys Glu Ile Asp  
1490 1495 1500

Val Lys Tyr Ile Gln Pro Gly Val Ala Ser Val Glu Glu Val Ile Glu  
1505 1510 1515 1520

Ala Lys Arg Val Leu Glu Lys Val Lys Asp Leu Ser Asp Glu Glu Arg  
1525 1530 1535

Glu Thr Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Val Glu Pro  
1540 1545 1550

Asn Asn Ala Ile Thr Val Asn Thr Gln Asn Glu Phe Thr Thr Lys Pro  
1555 1560 1565

Ser Ser Gln Val Thr Ile Ser Glu Gly Lys Ala Cys Phe Ser Ser Gly  
1570 1575 1580

Asn Gly Ala Arg Val Cys Thr Asn Val Ala Asp Asp Gly Gln Gln Pro  
1585 1590 1595 1600

<210> 11

<211> 29

<212> PRT

<213> Haemophilus influenzae

<400> 11

Val Asp Glu Val Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp  
1 5 10 15

Leu Ser Asp Glu Arg Glu Ala Leu Ala Lys Leu Gly  
20 25